



The Vertebrate TLR Supergene Family Evolved Dynamically by Gene Gain/Loss and Positive Selection Revealing a Host-Pathogen Arms Race in Birds

Khan, Imran; Maldonado, Emanuel; Silva, Liliana; Almeida, Daniela; Johnson, Warren E.; O'Brien, Stephen J.; Zhang, Guojie; Jarvis, Erich D.; Gilbert, M. Thomas P.; Antunes, Agostinho

Published in:
Diversity

DOI:
[10.3390/d11080131](https://doi.org/10.3390/d11080131)

Publication date:
2019





Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY](#)

Citation for published version (APA):
Khan, I., Maldonado, E., Silva, L., Almeida, D., Johnson, W. E., O'Brien, S. J., Zhang, G., Jarvis, E. D., Gilbert, M. T. P., & Antunes, A. (2019). The Vertebrate TLR Supergene Family Evolved Dynamically by Gene Gain/Loss and Positive Selection Revealing a Host-Pathogen Arms Race in Birds. *Diversity*, 11(8), [131].
<https://doi.org/10.3390/d11080131>

Article

The Vertebrate TLR Supergene Family Evolved Dynamically by Gene Gain/Loss and Positive Selection Revealing a Host–Pathogen Arms Race in Birds

Imran Khan ^{1,2}, Emanuel Maldonado ¹ , Liliana Silva ^{1,2}, Daniela Almeida ¹ , Warren E. Johnson ^{3,4}, Stephen J. O'Brien ^{5,6} , Guojie Zhang ^{7,8,9}, Erich D. Jarvis ^{10,11}, M. Thomas P. Gilbert ¹² and Agostinho Antunes ^{1,2,*} 

¹ CIIMAR/CIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, 4450-208 Porto, Portugal

² Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal

³ Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Road, Front Royal, VA 22630, USA

⁴ The Walter Reed Biosystematics Unit, Museum Support Center MRC-534, Smithsonian Institution, 4210 Silver Hill Rd. Suitland, MD 20746, USA

⁵ Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg 199004, Russia

⁶ Guy Harvey Oceanographic Center, Halmos College of Natural Sciences and Oceanography, Nova Southeastern University, 8000 N. Ocean Drive, Ft Lauderdale, FL 33004, USA

⁷ Section for Ecology and Evolution, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark

⁸ China National GeneBank, BGI-Shenzhen, Shenzhen 518083, China

⁹ State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

¹⁰ Laboratory of Neurogenetics of Language, Box 54, Rockefeller University, New York, NY 10065, USA

¹¹ Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

¹² Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark

* Correspondence: aantunes@ciimar.up.pt

Received: 19 June 2019; Accepted: 5 August 2019; Published: 12 August 2019



Abstract: The vertebrate toll-like receptor (TLRs) supergene family is a first-line immune defense against viral and non-viral pathogens. Here, comparative evolutionary-genomics of 79 vertebrate species (8 mammals, 48 birds, 11 reptiles, 1 amphibian, and 11 fishes) revealed differential gain/loss of 26 TLRs, including 6 (TLR3, TLR7, TLR8, TLR14, TLR21, and TLR22) that originated early in vertebrate evolution before the diversification of Agnatha and Gnathostomata. Subsequent dynamic gene gain/loss led to lineage-specific diversification with TLR repertoires ranging from 8 subfamilies in birds to 20 in fishes. Lineage-specific loss of TLR8-9 and TLR13 in birds and gains of TLR6 and TLR10-12 in mammals and TLR19-20 and TLR23-27 in fishes. Among avian species, 5–10% of the sites were under positive selection (PS) (omega 1.5–2.5) with radical amino-acid changes likely affecting TLR structure/functionality. In non-viral TLR4 the 20 PS sites (posterior probability PP > 0.99) likely increased ability to cope with diversified ligands (e.g., lipopolysaccharide and lipoteichoic). For viral TLR7, 23 PS sites (PP > 0.99) possibly improved recognition of highly variable viral ssRNAs. Rapid evolution of the TLR supergene family reflects the host–pathogen arms race and the coevolution of ligands/receptors, which follows the premise that birds have been important vectors of zoonotic pathogens and reservoirs for viruses.

Keywords: gene gain; gene loss; vertebrates; toll-like receptors; immune response; host–pathogen; positive selection

1. Introduction

The toll-like receptor (TLRs) supergene family members are type-I transmembrane glycoproteins belonging to the pathogen recognition receptor (PRRs) class of proteins expressed in the cell membrane and intracellular vesicles, including the endoplasmic reticulum, endosomes, lysosomes, and endolysosomes [1,2]. TLRs constitute one of the first lines of the immune defense system and recognize a variety of pathogen-associated-molecular patterns (PAMPs) during pathogen invasion, triggering the cascade of signaling pathways leading to the adaptive immune response [3–5]. The vertebrate TLRs studied to date are involved solely in immune response, in contrast with invertebrate TLR-like proteins, which are also associated with developmental functions [6].

The diversity of TLRs has facilitated the detection of a diverse group of pathogenic ligands [7]. The diverse mechanisms used by TLRs paralogs for TLR-ligand recognition and the formation of a m-shaped homo or heterodimeric complex, lead to activation of downstream signaling cascades by Toll/interleukin-1 receptor (TIR) domains [1,8–15]. The TIR dimer is recognized by the TIR domain present in different signaling adaptor proteins—including MyD88, MAL, TRIF, and TRAM [16]—and resulting in the activation of NFκB, the expression of various inflammatory and anti-pathogenic proteins [17], the initiation of the adaptive immune response, and the elimination of the invading pathogens [5,18–20].

Each TLR gene consists of a highly conserved intracellular (cytoplasmic) TIR domain, which is responsible for signal transduction [21], a conserved single transmembrane region and a variable extracellular domain (ECD), involved in the ligand recognition and dimerization. ECD consists of variable numbers (~16 to 28) of leucine-rich repeats (LRRs) motifs [22].

TLRs identified to date are grouped phylogenetically into six major families (TLR1, TLR3, TLR4, TLR5, TLR7, and TLR11) [23]. Gene gain and loss have been prominent features in the evolution of this supergene family. The large and variable number of TLR genes facilitate recognition of a wide variety of ligands from diverse pathogens (bacterial, fungal, protozoan, and virus). For example, recent studies have revealed important compensatory mechanisms of the TLR supergene family in both adaptive and innate response in the absence of major histocompatibility complex (MHC) II, CD4 and invariant chain (Ii) in cod fish [24], where TLRs act as an alternative to MHCII, a well-known conserved feature of the adaptive immune system of jawed vertebrates [24–26]. The innate immune receptors of the coelacanth are a mixture of mammalian- and teleost-specific TLRs, affirming the transitional position of coelacanth [27], evolutionarily connecting fish and tetrapods, along with its unique immune system lacking IgM [28].

We performed comparative evolutionary genomic analyses of vertebrate TLRs using whole genome sequencing data from 79 species (mammals, birds, reptiles, amphibians, and fishes) to understand the dynamics of gene gain and loss in shaping the TLR gene family repertoire in vertebrate genomes. The most important immunological function of TLRs comprehends protection from pathogens that requires the genes to evolve rapidly in response to selective pressure exerted by constantly evolving pathogens. Because avian species are often important vectors of zoonotic pathogens and reservoirs for viruses we analyzed avian TLRs (viral and non-viral TLRs) to determine how positive diversifying selection affecting the functional and structural variation of TLRs has improved pathogen recognition in birds. We discuss our findings in the context of host–pathogen interactions and the adaptive evolution of avian TLRs to diverse ecological conditions and adaptive requirements.

2. Materials and Methods

2.1. Genome Scan and Synteny Analysis of TLR Supergene Family

We used representative sequences of the vertebrate TLR1-27 as the query to exhaustively perform blast [29] searches to retrieve all available TLRs from 79 species, including the genomes of 48 birds [30,31] plus 31 other vertebrates which included 11 reptiles, a frog, and 11 fish genomes (Supplementary File 1), providing updated TLR information across vertebrate. TLRs, coding sequences retrieved from the Avian genome consortium [30,31] were used in subsequent adaptive evolution analysis. For the synteny analysis of TLRs we scanned vertebrate genomes (Supplementary File 2) for each TLR gene using extensive Blast searches [29] complemented with the Genomicus tool [32,33]. Gene loss can be hard to ascertain due to constraints imposed by levels of genome coverage [34]. Here we define gene loss as genes that are not found in a given assembly after extensive blast searches.

2.2. Phylogenetic Analysis

For the vertebrate TLR supergene family tree, sequences were aligned using ClustalW as implemented in MEGA5 [35] and were corrected manually. The tree was estimated with MEGA5 [35] software using the neighbor joining method with 1000 bootstrap replications. For the phylogenetic and adaptive analyses of the avian TLRs, one to one orthologs were aligned using Muscle [36], implemented in Seaview [37], and manually corrected. We chose the best substitution model based on the Akaike information criteria (AIC) in jModelTest 2 [38]. The sequences were used for maximum likelihood tree construction using PhyML [39] with 500 bootstrap replicates to check the robustness and reliability of the tree [40]. Sequences were tested for nucleotide substitution saturation using DAMBE 5 [41] by plotting the number of transition-transversion against the genetic distance using the F84 model [42], which allows for different equilibrium base frequency and transition-transversion rate bias for nucleotide substitution. The Xia test [43] implemented in DAMBE5 [41] was used to compare the index score (ISS) with critical score (ISS.C) at third and other codon positions to obtain estimates of saturation.

2.3. Gene Conversion and Recombination

The sequence alignments were tested for recombination using the online version of GARD (Genetic Algorithm for Recombination Detection) [44] available at <http://www.datamonkey.org>. The GENE-CONV [45] software was used for the detection of gene conversion events with 1000 permutation and Bonferroni corrected p -value cutoff of $p < 0.01$ and mismatch allowed ($/g1 = 1$).

2.4. Positive Selection in Avian TLRs

The important immunological function of TLRs, i.e., providing protection of host from pathogens, requires them to evolve rapidly in response to selective pressure exerted by rapidly and constantly evolving pathogens. In proteins, different functional sites experience distinct selection pressures with advantageous changes being positively selected due to adaptive benefits. The Avian TLRs are assumed to have similar ligands as those reported in mammalian TLRs [46,47]. In this study, we used multiple approaches to find signals of positive selection (i.e., $\omega > 1$, assessed by comparing the dN-number of non-synonymous substitutions per non-synonymous sites with that of dS-synonymous substitutions per synonymous sites) in avian TLRs (TLR1A, TLR1B, TLR2A, TLR2B, TLR3, TLR4, TLR5, TLR7, TLR15). We excluded TLR21 from positive selection analysis as insufficient sequences were found. The avian TLR7 is reported to have been duplicated in a few passerine birds [48–50] while TLR5 is pseudogenized in others [46]. Therefore, the TLR5 pseudogene and duplicated TLR7 sequences were excluded from the positive selection analysis.

We employed codon models implemented in PAML [51–53] and Datamonkey [54] together with amino acid models in TreeSAAP [55]. CODEML in PAML [51] implements likelihood ratio tests (LRT) for comparisons of sophisticated nested site-specific models calculated as twice the difference

of the log likelihood between the two models following Chi square distribution with the degrees of freedom corresponding to the difference in the number of parameters between the nested model, i.e., null model and alternate model (positive selection). A significant LRT implies that the null model is rejected and sites are under positive selection. We compared M1a (nearly neutral) vs. M2a (positive selection), and M7 (beta: assumes a beta distribution for $0 \leq \omega \leq 1$; therefore this model does not allow for positively selected sites) vs. M8 (beta& ω : same as the M7 plus an extra class of $\omega > 1$, allowing positive selection) to find sites under positive selection. Bayes empirical Bayes (BEB) inferred the posterior probabilities of positive selected sites where higher PP mean high confidence. We also used the Hyphy package (<http://www.hyphy.org>) [56] (<http://www.datamonkey.org/>) [54,57], which provides several approaches for detection of positive selected sites, including single-likelihood ancestral counting (SLAC), fixed effects likelihood (FEL), random effects likelihood (REL) [58], mixed effects model of evolution (MEME) [59], fast unconstrained Bayesian approximation (FUBAR) [60], and an integrative approach. The SLAC model uses ancestral sequences reconstruction, FEL calculates site-by-site dN/dS without assuming a prior distribution. REL assumes a prior distribution across site, FUBAR ensures robustness against model misspecification, and MEME is most appropriate to detect episodic diversifying selection affecting individual codon sites. In addition, the integrative approach incorporates all sites detected by SLAC, FEL, REL, FUBAR, and MEME. Sites detected by two different methods are interpreted as having more support of positive selection [61–63].

Further support for our results was gained by the complementary protein level approach implemented in TreeSAAP [55]. It uses ancestral sequence reconstruction to detect the physiochemical properties changes of the amino acid replacements considering 31 amino acid properties. The amino acid replacement can lead to conservative or radical change in physiochemical properties. The positive radical changes can modify the protein structure and/or function and the number of radical changes at a site can be used as an indicator of the positive selection strength. To facilitate interpretation of the level of changes at a site we compared sites having six or more radical changes (defined as type I) with sites with less than six properties (defined as type II).

2.5. Comparison of Domain Architecture, Homology Modeling, and Structure Analysis across Vertebrate TLRs

The LRRfinder was used to predict the domain architecture and define locations of specific amino acid residues in the TLR protein domain. This was also verified, whenever possible using the Uniprot protein database (<http://www.uniprot.org>). The structure of each TLR was predicted using the CPHmodels 3.2 protein homology modeling server, which resulted in significant model structure for the complete regions of TLR5 and the ECD region of TLR1A, TLR2A, TLR2B, TLR3, TLR4, and TLR7. No significant structure was predicted for TLR1B and TLR15. All the highly significant positive selected sites were displayed on the respective predicted structures to show the potential functional or structural significance of specific amino acid residues. The residues were mapped onto the predicted structure using PyMOL (<http://www.pymol.org>).

3. Results

3.1. Dynamic Gene Gain and Loss Shapes Vertebrate TLR Supergene Family Repertoire

The dynamics of gene gain and loss have an important role in the evolution and diversification of gene families [64–71]. These lineage and species-specific changes are important evolutionary mechanisms of adaptation [72–74]. Comprehensive TLR data from varied vertebrate species/lineages is crucial for assessing the major evolutionary events influencing the diversification of the vertebrate TLR supergene family. To describe and interpret TLR gene family evolution across vertebrates, we assessed representatives from major vertebrate groups by comparing closely and distantly related species (e.g., in mammals we covered monotremes-egg laying platypus, marsupial-short-tailed opossum, to placental mammals). TLR data was collected from diverse vertebrates. These included birds [30,31] and eight divergent reptiles from diverse environments (aquatic, semi-aquatic, and terrestrial) and under

unique adaptive pressures (pathogens), including three lizards (*Anolis carolinensis*, *Pogona vitticeps*, and *Gekko Japonicus*), two snakes (*Python molurus bivittatus* and *Protobothrops mucrosquamatus*), three turtles (*Pelodiscus sinensis*, *Chrysemys picta bellii*, and *Chelonia mydas*), and three crocodilians (*Alligator mississippiensis*, *Gavialis gangeticus*, and *Crocodylus porosus*). Together with mammals, amphibian, and fishes (see Table 1), our analyses revealed a comprehensive picture of the TLR supergene family evolution (Figures 1 and 2).

The fishes had the highest number of TLR subfamilies (20 subfamilies) with only six subfamilies (TLR6, TLR10–TLR12, TLR15, and TLR16) missing in fishes. This showed that most of the TLR subfamilies originated early in the fish lineage. Fishes also had the highest number of genes derived from duplication events (Table 1 and Figure 2), with cod having 41 TLR gene copies. These patterns highlight the importance of differential gene gain, and its prominence among fish lineages (Table 1 and Figure 2).

The comparative genomics of the TLR supergene family showed lineage and species-specific distribution—e.g., TLR16 is frog-specific, TLR24 lamprey-specific, TLR26 catfish-specific, and TLR27 coelacanth-specific—whereas TLR11–12 are present only in mammals and TLR19–20 and TLR26 are only found in teleost fishes. The TLR18 and TLR22 were lost in both the avian and mammalian lineages, suggesting similar adaptive pressure (Table 1 and Figure 2). Loss of TLR15 and TLR21 in the mammalian lineage possibly occurred after the divergence of sauropsida from the synapsida lineage (Table 1 and Figure 2). Likewise, the finding of TLR13, TLR18, and TLR22, together with TLR15 and TLR21 in reptiles (Table 1) suggests that these TLRs are not limited to particular vertebrate species and lineages, as previously suggested [23,45]. Our study reveals that dynamic gene gain and loss of TLRs in vertebrate genomes facilitated vertebrate TLR supergene family evolution under diverse ecological requirements.

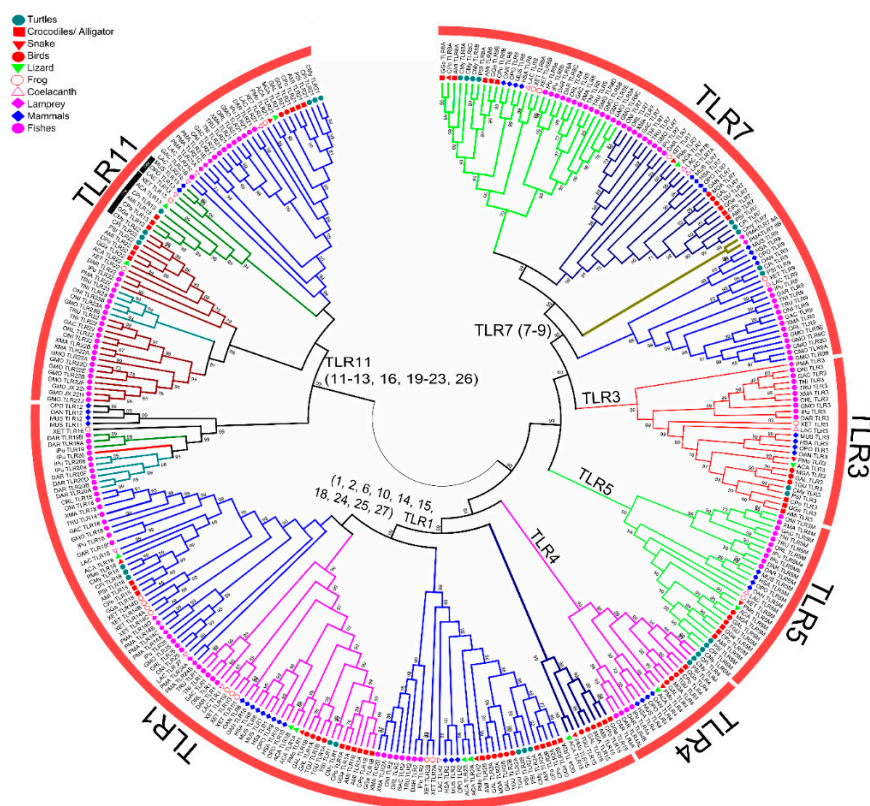


Figure 1. Phylogeny of vertebrate TLR supergene family from representative species (see Table 1). The NJ tree was made in MEGA5 with 1000 bootstrap replication.

Table 1. The TLR supergene family repertoire, showing gene gain and loss in different vertebrate species.

Class	Species Number	Species Scientific Name	TLR1										TLR3			TLR4	TLR5			TLR7			TLR11										Total F/P (Copies F/P)
			1	6	10	2	14	15	18	24	25	27	3	4	5	7	8	9	11	12	13	16	19	20	21	22	23	26					
Mammalia	1	<i>Homo sapiens</i>	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0(1)	0	0	0	0	0	0	0	10/1(10/1)					
	2	<i>Mus musculus</i>	1	1	0 (1)	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	12/1(12/1)					
	3	<i>Monodelphis domestica</i>	1	0	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	11/0(11/0)					
	4	<i>Ornithorhynchus anatinus</i>	1	0	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	11/0(11/0)					
	5	<i>Myotis lucifugus</i>	1	1	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	12/0(12/0)					
	6	<i>Equus caballus</i>	1	1	1	1	0	0	0	0	0	0	1	2	1	1	1	1	1	1	1	0	0	0	0	0	0	13/0(14/0)					
	7	<i>Canis familiaris</i>	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	10/0(10/0)					
	8	<i>Bos taurus</i>	1	1	1	1	0	0	0	0	0	0	1	1	1	1	0(1)	1	0	0	0	0	0	0	0	0	0	10/1(10/1)					
Aves	9	<i>Gallus gallus</i>	2	0	0	2	0	1	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	8/0(10/0)					
	10	<i>Meleagris gallopavo</i>	2	0	0	2	0	1	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	8/0(10/0)					
	11	<i>Taeniopygia guttata</i>	2	0	0	2	0	1	0	0	0	0	1	1	1	1	2	0	0	0	0	0	0	0	1	0	0	8/0(11/0)					
Reptilia	12	<i>Crocodylus porosus</i>	2	0	0	2	0	1	1	0	0	0	1	1	1	1	2	0	0	0	1	0	0	0	1	1	0	12/0(15/0)					
	13	<i>Gavialis gangeticus</i>	2	0	0	2	0	1	1	0	0	0	1	1	1	1	2	0	0	0	1	0	0	0	1	1	0	12/0(15/0)					
	14	<i>Alligator mississippiensis</i>	2	0	0	2	0	1	1	0	0	0	1	1	2	1	2	0	0	0	2	0	0	0	1	1	0	12/0(17/0)					
	15	<i>Pelodiscus sinensis</i>	2	0	0	2(1)	0	0	1	0	0	0	1	1	2	1	3	1	0	0	0	0	0	0	1	2	0	11/0(18/1)					
	16	<i>Chelonia mydas</i>	1	0	0	2(2)	0	0	1	0	0	0	1	1	2	1	2(1)	1	0	0	0	0	0	0	1	1	0	11/0(17/2)					
	17	<i>Chrysemys picta</i>	2	0	0	2(1)	0	0	1	0	0	0	1	1	2	1	3	1	0	0	1	0	0	0	1	1	0	12/0(18/1)					
	18	<i>Anolis carolinensis</i>	2	0	0	2	0	1	1	0	0	0	1	1	2	1	0	0	0	0	1	0	0	0	1	1	0	11/0(14/0)					
	19	<i>Python molurus</i>	1	0	0	2	0	1	1	0	0	0	1	1	2	1	0	0	0	0	0	0	0	0	1	1	0	10/0(12/0)					
	20	<i>Protobothrops mucrosquamatus</i>	1	0	0	3	0	1	0	0	0	0	1	1	2	1	0	0	0	0	1	0	0	0	1	0	0	9/0(12/0)					
	21	<i>Pogona vitticeps</i>	2	0	0	2	0	0	1	0	0	0	1	1	4	1	0	0	0	0	1	0	0	0	1	1	0	10/0(15/0)					
	22	<i>Gekko japonicus</i>	2	0	0	1	0	1	1	0	0	0	1	1	2	1	0	0	0	0	1	0	0	0	1	1	0	11/0(13/0)					
Amphibia	23	<i>Xenopus tropicalis</i>	2	0	0	2	4	0	0	0	0	0	1	0	2	1	2	1	0	0	1	1	0	0	1	1	0	12/0(20/0)					
Sarcopterygii (lobe-finned fishes)	24	<i>Latimeria chalumnae</i>	1	0	0	2	0	0	1	0	0	1	1	0	1(1)	2	1	1	0	0	1	0	0	0	3	0	0	11/0(16/1)					
Actinopterygii (ray-finned fishes)	25	<i>Xiphophorus maculatus</i>	1	0	0	2	0	0	1	0	0	0	1	0	2	1	1	1	0	0	0	0	0	0	1	3	6	11/0(20/0)					
	26	<i>Oryzias latipes</i>	1	0	0	1	0	0	1	0	1	0	1	0	2	1	1	1	0	0	0	0	0	0	1	1	0	11/0(12/0)					
	27	<i>Oreochromis niloticus</i>	1	0	0	5	0	0	1	0	1	0	1	0	2	1	2	1	0	0	0	0	0	0	3	3	3	12/0(24/0)					
	28	<i>Gasterosteus aculeatus</i>	1	0	0	1	0	0	1	0	0	0	1	0	3	1	1	1	0	0	0	0	0	0	2	1	0	10/0(13/0)					
	29	<i>Takifugu rubripes</i>	1	0	0	1	0	0	1	0	0	0	1	0	2	1	1	1	0	0	0	0	0	0	1	1	1	11/0(12/0)					
	30	<i>Tetraodon nigroviridis</i>	1	0	0	1	0	0	1	0	0	0	1	0	2	1	1	1	0	0	0	0	0	0	1	1	1	11/0(12/0)					
	31	<i>Gadus morhua</i>	0	0	0	0	0	0	1	0	7	0	1	0	0	1	12	5	0	0	0	0	0	0	1	12	1	9/0(41/0)					
	32	<i>Danio rerio</i>	1	0	0	1	0	0	1	0	0	0	1	3	2	1(2)	3	1	0	0	0	0	2	6	1	1	0	13/0(26/2)					
	33	<i>Ictalurus punctatus</i>	1	0	0	1	0	0	1	0	1	0	1	2	3	1	2	1	0	0	0	0	1	2	1	1	0	15/0(20/0)					
Cephalaspidomorphi (lampreys)	34	<i>Petromyzon marinus</i>	0	0	0	0	4	0	0	2	0	0	1	0	0	7/8	7/8	0	0	0	0	0	0	3	1	0	6/0(13/0)						

The color indicates gene numbers. The absence is shown in yellow with “0”. Presence without duplicates in red with “1” and number of duplicates copies in blue. The “F/P” indicates functional/pseudogene TLRs.

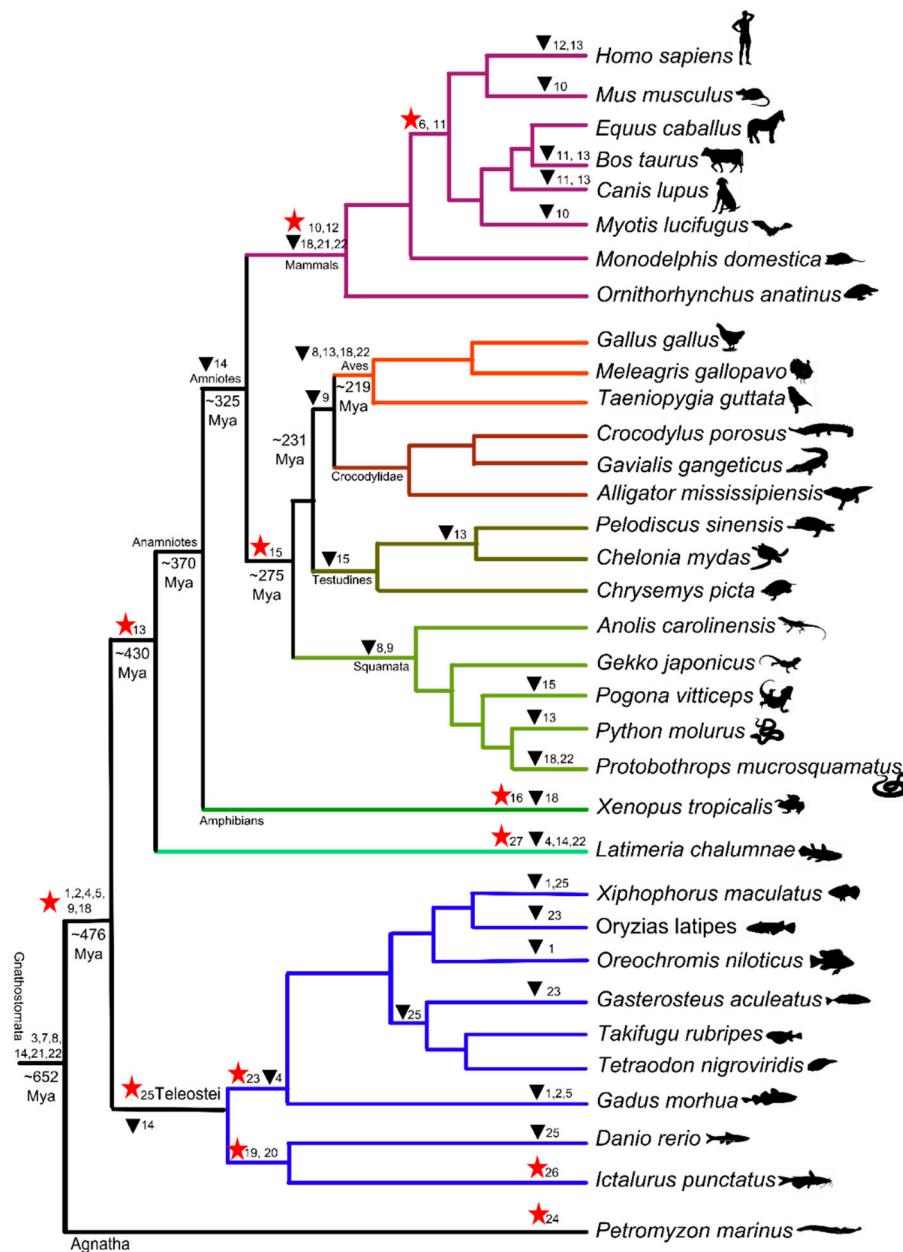


Figure 2. Molecular evolution of vertebrate TLR gene family with events of TLR subfamily gains and losses shown on the consensus phylogeny made based on published literature. The dates were inferred using <http://www.timetree.org/>. The gains are represented by a star and the triangle shows loss of TLR subfamily.

3.2. Synteny of TLR Supergene Family in Vertebrates

Synteny analysis has been fundamental in identifying conserved genomic arrangements and orthologous relationships. Synteny analyses highlighted the genomic organization of TLRs (Figure 3, Supplementary File 2 A-P) which helped evaluate and interpret TLR gene family evolution and levels of homology of TLR genes in different genomes. Lineage-specific TLRs, e.g., fish-specific and tetrapod-specific TLRs, showed highly conserved synteny. In contrast, the coelacanth showed shared features with both fish and tetrapods. Only TLR3, TLR5M, TLR7, and TLR8 genes showed conserved synteny across vertebrates (from fishes to mammals) (Figure 3).

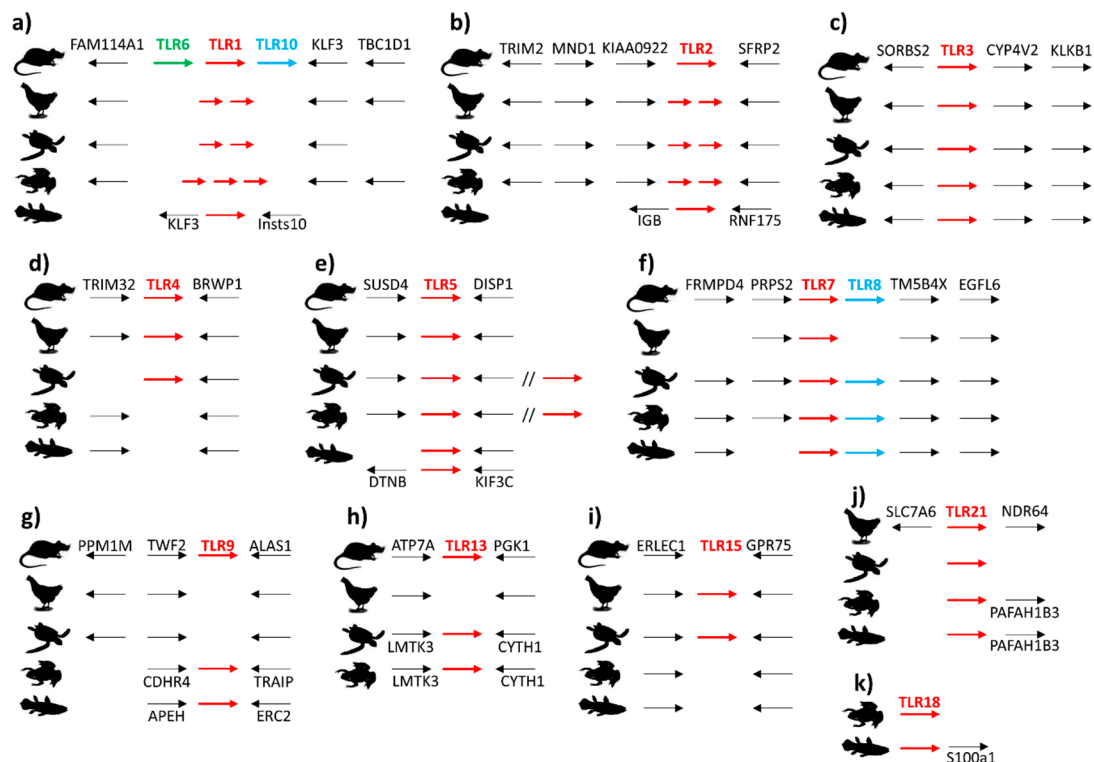


Figure 3. Genomic organization of TLRs clusters present in two or more classes of vertebrates. (a) Syntenic organization of TLR6, TLR1, and TLR10 genes whereas only TLR1 is widespread across all vertebrates; (b) Syntenic organization of TLR2 genes; (c) Syntenic organization of TLR3; (d) Syntenic organization of TLR4 only present in mammals, avians, and reptiles; (e) Syntenic organization of TLR5; (f) Syntenic organization of TLR7 and TLR8; (g) Syntenic organization of TLR9 which is absent in avian and reptilian; (h) Syntenic organization of TLR13 present in tetrapods (with exception of avian); (i) Syntenic organization of TLR15 exclusive of reptiles and avian; (j) Syntenic organization of TLR21.

3.3. Phylogenetic Analysis of Vertebrate TLR Supergene Family

The TLR data from diverse vertebrates (Figure 1 and Table 1) used for the phylogenetic reconstruction of the TLR supergene family strongly supports the classification of TLRs into six families [23], along with new subfamilies that shape the dynamic distribution of the TLR supergene family (Table 1 and Figure 1). Most of the new subfamilies belong to TLR family 1 and 11, as 10 subfamilies TLR1, TLR2, TLR6, TLR10, TLR14, TLR15, TLR18, TLR24, TLR25, and TLR27 grouped in TLR1, and 10 subfamilies TLR11–TLR13, TLR16, TLR19–TLR23, and TLR26 grouped in TLR11. In contrast, no new subfamilies gain in the remaining TLR families TLR3, TLR4, TLR5, and family TLR7 were recognized (subfamilies TLR7, TLR8, and TLR9).

We found that the TLR13 clade consists of two diverse groups, one with coelacanth and mammals (labeled as 13M) and another with amphibian and reptiles (labeled as 13X) (Figure 1), which were supported by high bootstrap values and our synteny analysis (Supplementary File 2 H1–H2). We clarified the dubious naming of TLR14, TLR18, and TLR25 and address this ambiguity by assessing synteny (genomic arrangement of gene and flanking genes) and relative phylogenetic relationships. Our phylogeny and syntenic analyses suggest that TLR14 [75], which was first described in *Fugu*, should also be placed within TLR18 with zebrafish and catfish and be renamed TLR18 (Supplementary File 2–I2). The genes that have been previously referred to as TLR18 in medaka, tilapia, and cod are syntenic and form a monophyletic clade with TLR25 from catfish [76] and should thus be classified as TLR25 (Figure 1 and Supplementary File 2P).

Since the frog TLR12 is more similar to the zebrafish TLR19 than the mammalian TLR12, and does not retain a conserved syntenic relationship with either TLR12 and TLR19, we retained it as

TLR16 [23]. Overall, this TLR dataset, derived from an extended evaluation of additional newly sequenced genomes, provided an improved evolutionary assessment and a more accurate classification of the TLR supergene family phylogeny across vertebrates (Table 1, Figure 2).

Gene Conversion and Recombination

Gene conversion is common in tandem-duplicated gene paralogs located in close proximity. Gene conversion and recombination can bias phylogenetic inference and therefore care should be taken to identify signals of those events. The duplicated avian TLR gene paralogs of TLR1 and TLR2 are closely related to their orthologous counterpart from other species (i.e., within species paralogs TLR1A is closely related to TLR1B and TLR2A with TLR2B). This is caused by concerted evolution, possibly through gene conversion, leading to homogenization of the paralogs and reduced phylogenetic signals. Gene conversions have been reported in avian TLR1 and TLR2 [77]. TLR1 and TLR2 are closely located and are separated by a short physical distance (~12 kb and ~5 Kb between TLR1 and TLR2 duplicates, respectively) [46,64]. The C terminal of TLR1A/B ranging from LRR14 to the TIR domain is under gene conversion, whereas in TLR2A/B, the N terminal region (N terminal to LRR8) and C terminal region (LRR15 till TIR) are homogenized [46,64,77]. These homogenized regions may have important conserved functions [77].

3.4. Vertebrate TLRs Domain Architecture

TLRs consist of three major characteristic domains (extracellular domain (ECD), transmembrane domain (TM), and toll/interleukin-I receptor domain (TIR)). The one exception is TLR5S, which lacks the characteristic TIR domain. The TM domain connects the ECD with the cytoplasmic domain. The ECD is a solenoid shaped structure involved in interaction with PAMPs present in varied pathogens and is also involved in the formation of a M shaped homo and/or hetero dimer, which leads to signaling cascade by TIR activation [78].

The comparison of predicted architecture of ECD from various lineages of vertebrates: stickleback, coelacanth, Chinese soft-shell turtle, lizard, chicken (Figure 4, Supplementary File 3) [78] showed variable numbers of leucine rich repeats (LRRs). We also compared the domain architecture of TLRs from stickleback, coelacanth, Chinese soft shell turtle, *Anolis* lizard, and chicken and found a single domain in TLR 3, 5, 7, 8, 9, 13, 15, 18, 21, 22, and 27, whereas three domains were present in TLR 1, 2, and 4 (except in the coelacanth TLR1 exhibiting just a single domain) (Figure 4, Supplementary File 3).

Each LRR was around 20–30 amino acids long and had a conserved LxxLxLxxN leucine rich motif and a remaining variable region. The hydrophobic leucine residues constitute the conserved concave surface of the parallel beta strands forming the hydrophobic core, where asparagine is involved in the hydrogen bonding providing structural integrity [79]. Leucine can be replaced with other hydrophobic amino acids, whereas asparagine can also be replaced with other hydrogen donor's like threonine, serine, and cysteine. The variable 'x' residues are responsible for the TLR function.

The exposed and convex surface of ECD is formed by the variable part of LRR repeat and this region is involved in PAMPs recognition. The cysteine clusters capping present in the terminal LRRs (LRR-NT and LRR-CT) protect the terminal hydrophobic residues. The human TLR1, 2, 4, 6, and 10 have all three domains, whereas TLR 3, 5, 7, 8, and 9 have a single domain [79]. This categorization is due to the interrupted asparagine in LRRs of the central domain, which results in structural flexibility, whereas uniform LRRs repeats with continuous asparagine results in single domain architecture [78]. The comparison of the domain architecture of representative TLRs from stickleback, coelacanth, Chinese soft shell turtle, *Anolis* lizard, and chicken was consistent with the human TLRs (Figure 4, Supplementary File 3).

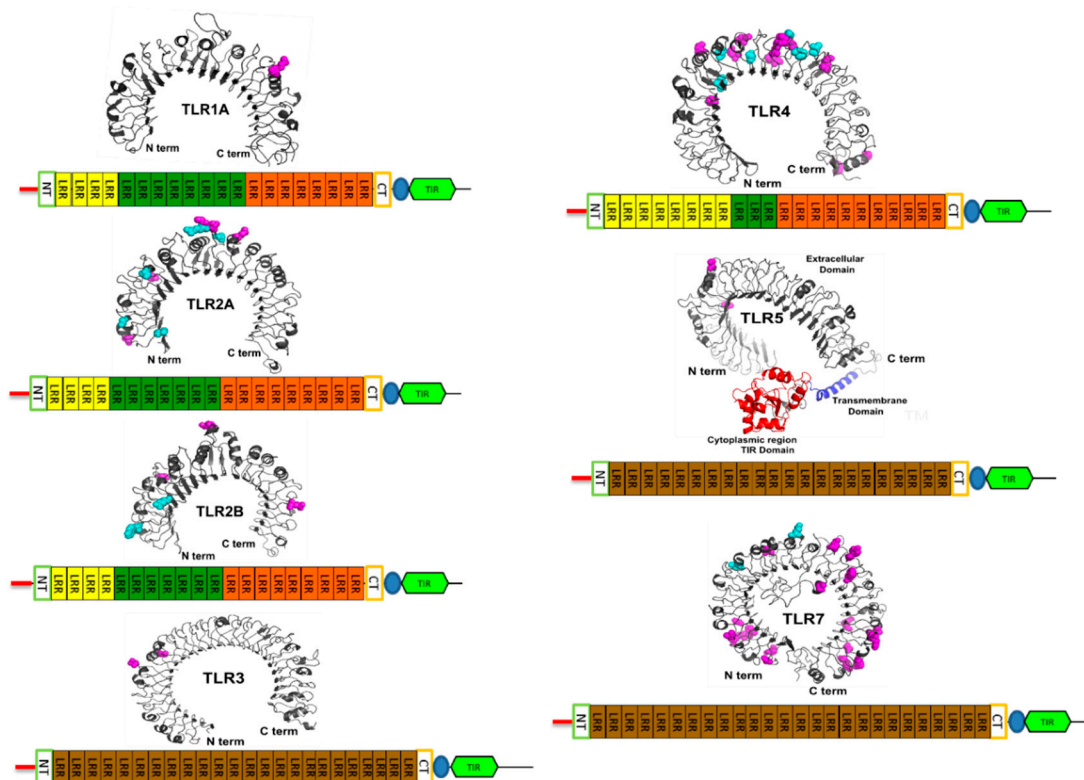


Figure 4. Scheme of TLR domain architecture for avian TLRs and the positive selected sites with PP > 0.99 in M8 which also had PP > 90 in M2a are shown in the predicted structure of respective TLRs with magenta color. The site with cyan color represents type I changes detected by TreeSAAP. The numbering of positive selected sites in as per sequences given in Supplementary File 4 Tables (For TLR1B and TLR15 structure prediction was not significant).

3.5. Rapid Adaptive Evolution of Avian TLRs

We found variable numbers of positive selection sites among different TLR genes. The positive selection model M2a and M8 implemented in PAML4.7 [51] detected significant signals of positive selection in all genes (Table 2). The reliability of sites was well supported by the fact that most positive selected sites with PP > 0.99 under M8 also had PP > 0.90 in M2a. The M8 model found higher number of sites compared to the more conservative model M2a (Table 2). We found significant evidence of positive selection in all avian TLRs studied with different percentage of positively selected sites for different genes ranging from 11% in TLR15 to 5% in TLR7. The highest omega values were found in the TLR4, TLR2A, and TLR7 with 2.6, 2.6, and 2.5, respectively, while TLR15 had the lowest value of 1.5.

The number of positive selected sites with the highest (BEB) posterior probabilities (PP) also varied and the maximum number of sites with high PP were found in TLR7 (viral) and TLR4 (non-viral), which shows for the first time that both viral and non-viral TLR genes follow a positive selective regime (Table 2). In TLR7, Model M8 detected high numbers of positive selected (PS) sites (4.5% sites, total of 42 positively selected sites, 30 sites PP > 0.90, 25 sites PP > 0.95 and 23 sites PP > 0.99) and a somewhat similar scenario was found in TLR4 (5.5% sites, total of 45 sites, 29 sites PP > 0.90, 27 sites PP > 0.95 and 20 sites PP > 0.99). TLR3 had fewer sites with high PP (e.g., of 5% positive selected sites in TLR3 only five sites had PP > 0.90). Overall these results show that the number and strength of positive selection sites varies in the TLR supergene family and that both viral (TLR7) and non-viral (TLR4) TLRs evolved under strong positive selection.

The avian TLR1/TLR2 forms heterodimers that are activated by both diacyl (Malp-2) and triacyl (Pam3) lipopeptides [13,79,80], with the exception of TLR2a/TLR1b, which are activated by Pam3 but not by Malp-2. In addition, TLR2a/TLR1b is activated by peptidoglycan [77]. We found 12 positively-selected (PS) sites in TLR1A, 26 PS sites in TLR1B, 34 PS sites in TLR2A, and 25 PS sites in

TLR2B. The previously reported PS site 304 in TLR2A [46,77] and PS sites 293, 295, and 296 found in TLR2B [46], corresponding to the chicken sequence, were also found in our study (Table 3).

Overall, of 12 sites in TLR1A (Tables 2–4, Figure 4, Supplementary File 4), 10 PS sites were present in LRRs, of which five were in the variable region and possibly involved in the PAMP binding. Only one site each was present in LRR-CT and TM. All 26 sites in TLR1B (Tables 2–4, Figure 4, Supplementary File 4) were located in ECD with 7 and 13 sites in conserved and variable regions, respectively, and 4 sites in LRR-NT and 2 in LRR-CT. In TLR2A, 2 of 34 sites (Tables 2–4, Supplementary File 4) were in the signal peptide, 4 in LRR-NT, and 17 of the remaining 28 sites present in LRR were in the variable region. All of the 34 sites found in TLR2A were present in ECD with 2 sites in the signal peptide region, 4 sites in LRR-NT, and the other 28 in LRRs (21 in variable regions and remaining 7 in conserved regions). We found 25 sites in TLR2B (Tables 2–4, Figure 4, Supplementary File 4), all of them in LRRs with 14 in the variable region and 11 in the conserved region. The PS sites found in our study could be related with the a wide range of ligands as TLR2 recognizes a variety of compounds other than triacyl lipoproteins, including lipoteichoic acids, lipoarabinomannan, and zymosan [18]. Also, the formation of combinatorial binding sites by selection of TLR1 or TLR6 as the dimerization partner can explain, at least in part, the broad ligand specificity. A similar mechanism might explain the extent that PS sites found in TLR1 and TLR2 are shared with their counterparts in birds. The comparison of the chicken TLR2A and TLR2B genes with respect to the human TLR2s reveals that PS sites 280, 292, 304, 308, 309, 311, 312, 315, 335, 344, 356, 372, 392, 393, and 413 of TLR2A and PS sites 260, 274, 295, 297, 298, 299, 302, 317, 328, 329, 343, 390, and 401 of TLR2B map at or near to the ligand-binding domain and the dimerization surface, as identified from the crystal structure of the complex TLR2–TLR1 with the tri-acylated Pam3CSK4 lipoprotein in humans and mice [13,81].

TLR3 recognizes dsRNA and prevents the spread of most viruses. TLR3-ECD binds with dsRNA at two sites located at opposite ends of the TLR3 horseshoe structure. The first dsRNA:TLR3 interaction site is located close to the C-terminus, on LRR19–LRR21 and the second dsRNA:TLR3 interaction site is located on the N-terminal end (LRR-NT–LRR3) [82]. The intermolecular contact between the two C-terminal domain regions of TLR3 coordinates and stabilizes the dimer by a series of protein–protein interactions. Of 22 PS sites in TLR3 (Tables 2–4, Figure 4, Supplementary File 4), 2 are in signal peptides, 2 in LRR-NT, 2 in TM domains, 3 in the TIR domain, and 13 in LRRs. Of 13 sites in LRRs, 9 were in the variable region. The PS sites found in TLR3 likely favor the recognition and protection against rapidly evolving viral RNA [82].

TLR4 forms a heterodimer with myeloid differentiation factor 2 (MD-2) and recognizes diverse LPS molecules [83,84] along with components of yeast, trypanosoma, and even viruses [20,81,85]. The ECD of TLR4 consists of three subdomains. N subdomain consists of LRR-NT and LRR 1–6, the central subdomain ranges from LRR7–12 and the C subdomain consists of LRR 13–22 and LRR-CT. LPS causes dimerization of the TLR4–MD-2 complex at the central and/or the C-terminal, and interaction between TLR4 and LPS–MD-2 complex takes place at the concave surface between N and central sub-domain [84]. We found 41 PS sites in TLR4 (Tables 2–4, Figure 4, Supplementary File 4), three which were present in TM domain and one in TIR domain. The remaining 37 sites are present in ECD, with 3 in LRR-CT and the others 34 in LRRs with 23 of these in the variable region and the 11 others in the conserved region. The majority of the PS sites are present in regions involved in ligand binding and dimerization.

The TLR4 has a 330-amino-acid-long variable region in ECD which is also known as the middle region. Within the middle region there are 82 amino acids that are hyper-variable across species with numerous species-specific changes [86]. The majority of the PS sites found in TLR4 were concentrated near to the hyper-variable region. The primary contact interface between TLR4 and MD-2 involves two chemically distinct regions, the A and B patches provided by the N-terminal and the central domains of TLR4 and main dimerization interface of TLR4 located in the central C-terminal domain. The presence of positive selected sites in TLR4–MD-2–LPS complex may support the remarkable versatility of the

ligand recognition mechanisms employed by the TLR family, which is essential for defense against diverse microbial infections [84,86].

The bacterial flagellin is a virulence factor recognized by TLR5 [87]. The residues 174–401 in ECD of TLR5 are responsible for species-specific flagellin recognition [88]. The alpha and epsilon Proteobacteria are able to evade TLR5 recognition by mutating key residues in the TLR5 recognition site [89] and it is suggested that positive selection in primate TLR5 may be related with coevolution between PRRs and their microbial ligands [81]. Twenty-six PS sites were found in TLR5 (Tables 2–4, Figure 4, Supplementary File 4), with two sites each in signal peptide, LRR-NT, TM and TIR domains, one site in LRR-CT and the remaining 17 sites in LRRs, of which seven were in the variable region and 10 in the conserved region. Seven of these PS sites were located in 228 amino acid region identified previously [88] and thus could play an important role in the species-specific flagellin recognition and defense.

TLR7 recognize single-stranded RNAs [90]. In TLR7, we found 4, 1 and 4 PS sites in LRR-NT, TM, and TIR domain, respectively, and 39 PS sites were present in LRRs of which 33 sites were in the variable region (Tables 2–4, Figure 4, Supplementary File 4). This variable region directly interacts with the ssRNA. Thus, the high proportion of PS sites in viral TLR7 is probably related with the host pathogen arms race. TLR7 is also known to be evolving under strong selection pressure in mammals [79,85]. Recent studies found TLR7 to be under positive selection in bats, which are natural reservoirs and carriers of numerous deadly viruses [91]. The long-term coexistence of bats and viruses imposed strong selective pressures on the bat genome, with strong positive selection occurring in genes like TLR7 involved in the innate immune system, the first line of anti-viral defense [92]. The rapid evolution of viral TLR and its associated role in maintaining and disseminating viruses (e.g., in bats), suggest that it might have a similar role in the rapid evolution of viral TLR7 in birds. The avian TLR7 is a candidate for the detection of influenza [93] and has evolved to recognize ssRNA (with very high mutation rates due to lack of mismatch repair), pointing towards the long coexistence of viruses and birds. This may help understand the role of birds as natural reservoirs and vectors of zoonotic pathogens [93].

TLR15s recognize yeast-derived agonists [94]. We found 41 PS sites (Table 3 and Supplementary File 4) with 1, 3, 4, 2, and 1 PS sites in the signal peptide, LRR-NT, LRR-CT, TM, and TIR domains, respectively, and 30 sites in LRRs, of which 14 were present in the variable region. None of the PS sites were located in the highly conserved three box regions of the TIR domain [95,96]. The exact mechanism of action of TLR15 is unknown but the presence of PS sites is consistent with other TLRs involved in immune defense.

Evidence of positive-selected sites was further assessed using by multiple complementary approaches (SLAC, FEL, REL, MEME, and FUBAR) implemented in the HyPhy package (Table 3) (<http://www.hyphy.org>) [56] (<http://www.datamonkey.org/>) [54,57] in viral and non-viral avian TLRs (Table 3). Sites detected by more than one method (concordant of two or more methods) were thus considered to be robust candidates for positive selection (Table 3) [79]. We further implemented TreeSAAP [55] to detect positive radical changes in physiochemical properties of amino acids, which in turn can affect the structure and function of proteins. The TreeSAAP results also complemented our findings of positive selection (Table 4 and Figure 4), with most sites under selection belonging to type I (sites with six or more positive radical changes) or type II radical changes (sites with less than six positive radical changes).

Finally, homology modeling elucidated the role of the highly significant positive selected sites (Table 4) on the structural and functional diversification of TLRs (Figure 4). Since the majority of positive selected sites restricted to the variable extracellular domain (ECD) are involved in the ligand recognition and dimerization, this highlights their importance in the structure and function of these proteins (Figure 4). This strongly supports the hypothesis that the host–pathogen arms race drives the rapid evolution of TLRs.

Table 2. PAML results for nested site model comparisons for test of positive selection.

Gene	Model	Likelihood (lnL)	Parameters	2ΔlnL (LRT)	Significance (P-Value)	No. of PS Sites
TLR1A	M1A	−8578.3621	p:0.65788 0.34212 w:0.07820 1.00000			
	M2A	−8568.3574	p:0.65059 0.30720 0.04221 w:0.08172 1.00000 2.43025	20.0096	4.52×10^{-5}	9, 1 *, 0 **
	M7	−8574.9757	p = 0.24899 q = 0.48700			
	M8	−8557.0925	p0 = 0.93283 p = 0.32098 q = 0.81643 (p1 = 0.06717) w = 1.97137	35.7664	1.71×10^{-8}	9, 5 *, 1 **
TLR1B	M1A	−15,198.7487	p:0.64884 0.35116 w:0.09042 1.00000			
	M2A	−15,173.1918	p:0.63406 0.31995 0.04599 w:0.09272 1.00000 2.46912	51.1138	7.96×10^{-12}	12, 7 *, 2 **
	M7	−15,185.5024	p = 0.25368 q = 0.45987			
	M8	−15,147.3293	p0 = 0.91288 p = 0.34184 q = 0.83943 (p1 = 0.08712) w = 1.88577	76.3463	2.64×10^{-17}	20, 14 *, 4 **
TLR2A	M1A	−14,595.1078	p:0.58122 0.41878 w:0.14194 1.00000			
	M2A	−14,510.3897	p:0.53089 0.40739 0.06172 w:0.13788 1.00000 3.10556	169.4363	1.61×10^{-37}	18, 13 *, 11 **
	M7	−14,590.6910	p = 0.34904 q = 0.40071			
	M8	−14,500.5430	p0 = 0.91485 p = 0.40926 q = 0.52204 (p1 = 0.08515) w = 2.56498	180.2960	7.07×10^{-40}	20, 16 *, 11 **
TLR2B	M1A	−17,453.3413	p:0.70636 0.29364 w:0.08820 1.00000			
	M2A	−17,429.5429	p:0.69833 0.27518 0.02650 w:0.09079 1.00000 2.72661	47.5967	4.62×10^{-11}	10, 6 *, 0 **
	M7	−17,439.5787	p = 0.24244 q = 0.52859			
	M8	−17,403.8206	p0 = 0.94556 p = 0.29832 q = 0.80619 (p1 = 0.05444) w = 1.98273	71.5161	2.95×10^{-16}	12, 10 *, 5 **
TLR3	M1A	−17,682.7154	p:0.70932 0.29068 w:0.10081 1.00000			
	M2A	−17,667.5931	p:0.70389 0.28584 0.01027 w:0.10207 1.00000 3.25211	30.2446	2.71×10^{-7}	2, 2 *, 2 **
	M7	−17,680.6455	p = 0.29425 q = 0.62674			
	M8	−17,654.9577	p0 = 0.94774 p = 0.38756 q = 1.03153 (p1 = 0.05226) w = 1.87953	51.3755	6.98×10^{-12}	5, 2 *, 2 **
TLR4	M1A	−27,142.6224	p:0.65430 0.34570 w:0.08873 1.00000			
	M2A	−26,989.6674	p:0.63053 0.32499 0.04449 w:0.08946 1.00000 3.12384	305.9101	3.74×10^{-67}	24, 19 *, 16 **
	M7	−27,104.3280	p = 0.24737 q = 0.45214			
	M8	−26,948.5913	p0 = 0.94429 p = 0.28421 q = 0.58752 (p1 = 0.05571) w = 2.60768	311.4734	2.31×10^{-68}	29, 27 *, 20 **

Table 2. Cont.

Gene	Model	Likelihood (lnL)	Parameters	2ΔlnL (LRT)	Significance (P-Value)	No. of PS Sites
TLR5	M1A	−10,761.5148	p:0.64778 0.35222 w:0.07521 1.00000			
	M2A	−10,754.4425	p:0.65397 0.29395 0.05208 w:0.08616 1.00000 2.27090	14.1446	8.48×10^{-4}	0, 0 *, 0 **
	M7	−10,767.1968	p = 0.16412 q = 0.25713			
	M8	−10,751.9386	p0 = 0.88630 p = 0.31653 q = 0.82154 (p1 = 0.11370) w = 1.85950	30.5162	2.36×10^{-7}	11, 2 *, 0 **
TLR7	M1A	−31,101.4634	p:0.72733 0.27267 w:0.06793 1.00000			
	M2A	−30,961.2534	p:0.71077 0.25222 0.03700 w:0.06917 1.00000 3.03247	280.4200	1.28×10^{-61}	25, 24 *, 19 **
	M7	−31,044.1197	p = 0.18953 q = 0.48287			
	M8	−30,894.4389	p0 = 0.95423 p = 0.22385 q = 0.67658 (p1 = 0.04577) w = 2.52048	299.3616	9.87×10^{-66}	30, 25 *, 23 **
TLR15	M1A	−38,867.7085	p:0.66864 0.33136 w:0.09826 1.00000			
	M2A	−38,835.1856	p:0.65901 0.31957 0.02142 w:0.09812 1.00000 2.29015	65.0457	7.51×10^{-15}	8, 8 *, 5 **
	M7	−38,651.1989	p = 0.32880 q = 0.73980			
	M8	−38,608.5273	p0 = 0.94189 p = 0.39709 q = 1.16306 (p1 = 0.05811) w = 1.50749	85.3434	2.94×10^{-19}	9, 8 *, 3 **

The number of positively selected sites are shown in increasing order of posterior probability with PP > 0.90%, PP > 0.95% and PP > 0.99%. The number of positively selected sites are shown in increasing order of posterior probability. Sites with PP > 0.90%, (*) PP > 0.95% and (**) PP > 0.99%. The numbering of positive selected sites in as per sequences given in Supplementary File 4 Tables S2–S10.

Table 3. Positive selected sites identified by various methods (PAML-M8, SLAC, FEL, REL, MEME, FUBAR, and Integrated).

Gene	M8 ^a	SLAC	FEL	REL	MEME	FUBAR	Integrated ^b	Total Common Sites ^c (X/Y)
TLR1A	258, 400, 408 , 423, 450, 460, 483, 502, 591	<u>388, 429, 566</u>	<u>294, 342, 384, 388, 429,</u> <u>460, 535, 566, 611</u>	<u>297, 301, 384, 388, 400,</u> <u>429, 438, 460, 463, 470,</u> 550, 566, 611	266, 284, <u>294, 346, 384,</u> <u>388, 400, 411, 429, 460,</u> <u>461, 463, 518, 535, 550,</u> 559, <u>566, 571, 599, 608,</u> <u>611</u>	<u>388, 429, 460,</u> <u>566</u>	266, 284, <u>294, 297, 301, 342, 346,</u> <u>384, 388, 400, 408, 411, 429, 438,</u> <u>460, 461, 463, 470, 518, 535, 550,</u> 559, <u>566, 571, 599, 608, 611</u>	<u>11/12</u>
TLR1B	<u>41, 59, 88, 122,</u> <u>148, 149, 168,</u> <u>175, 248, 256,</u> <u>277, 286, 298,</u> <u>308, 318, 350,</u> 357, 439, 486, 488	<u>38, 41, 119, 120,</u> <u>122, 123, 148,</u> <u>168, 175, 216,</u> <u>232, 277, 398,</u> <u>414</u>	<u>11, 38, 41, 49, 62, 67,</u> <u>119, 120, 122, 123, 128,</u> <u>148, 156, 167, 168, 175,</u> <u>216, 232, 266, 277, 308,</u> <u>371, 398, 408, 414, 627</u>	<u>38, 41, 59, 119, 120, 122,</u> <u>123, 148, 150, 168, 175,</u> <u>181, 207, 216, 232, 248,</u> <u>277, 286, 308, 311, 398,</u> <u>414, 485</u>	26, 38, 41, 43, 44, 62, 92, <u>119, 120, 122, 123, 127,</u> <u>144, 146,</u> 148, 156, 167, 168, 175, 194, 206, 214, 216, 232, 266, 277, 308, 309, <u>311,</u> 353, 366, 371, 398, 407, 408, 414, 432, 447, 639	<u>38, 41, 119,</u> <u>120, 122, 123,</u> <u>148, 168, 175,</u> <u>216, 232, 277,</u> <u>308, 398, 414</u>	11, 26, 38, 41, 43, 44, 49, 59, 62, 67, 92, 119, 120, 122, <u>123, 127, 128, 144,</u> <u>146, 148, 150, 156, 167, 168, 175,</u> 181, 194, 206, 207, 214, 216, 232, 248, 266, 277, 286, 298, 308, 309, <u>311, 353, 366, 371, 398,</u> 407, <u>408, 414, 432, 447, 485, 627, 639</u>	<u>25/26</u>
TLR2A	<u>7, 16, 59, 67,</u> <u>108, 129, 171,</u> <u>206, 220, 270,</u> <u>304, 307,</u> <u>308, 311, 312,</u> <u>338, 363,</u> <u>372, 393, 413</u>	<u>16, 63, 108,</u> <u>292, 304, 308,</u> <u>311, 315, 335,</u> <u>393, 418, 419</u>	<u>16, 28, 44, 45, 63, 74,</u> <u>77, 108, 171, 209, 217,</u> <u>235, 257, 292, 304, 308,</u> <u>309, 311, 315, 335, 344,</u> <u>356, 372, 393, 418, 419</u>	<u>7, 16, 67, 108, 129, 171,</u> <u>217, 235, 276, 280, 292,</u> <u>306, 308, 311, 312, 335,</u> <u>356, 372, 387,</u> <u>392, 393, 413, 416, 418,</u> <u>419</u>	16, 19, 28, 44, 63, 67, 77, 96, 108, 122, 125, 138, 171, 174, 187, 195, <u>209,</u> 247, 250, <u>257, 264, 277,</u> <u>280, 292, 294,</u> 304, 308, 309, 311, 315, 322, 335, 344, 347, 349, 367, <u>372, 393, 418, 419,</u> <u>425</u>	<u>16, 108, 171,</u> <u>292, 308, 311,</u> <u>372, 392, 413, 418</u>	<u>7, 16, 19, 28, 44, 45, 59, 63, 67, 74,</u> <u>77, 96, 108, 122, 125, 129, 138, 171,</u> 174, 187, 195, 206, 209, 217, 235, 247, 250, 257, 264, 276, 277, <u>280,</u> <u>292, 294, 304, 306, 308,</u> 309, <u>311, 312, 315, 322, 335, 338,</u> <u>344, 347, 349, 356, 367, 372, 387,</u> <u>392, 393, 413, 416, 418, 419, 425</u>	<u>31/34</u>
TLR2B	50, 58, 99, 162, <u>175, 211, 260,</u> <u>295, 297, 298,</u> <u>329, 456</u>	<u>89, 99, 137,</u> <u>162, 176, 297,</u> <u>328, 329, 390</u>	<u>42, 48,</u> <u>89, 99, 137, 149, 162, 176,</u> <u>199, 200, 219, 257,</u> <u>274, 295, 297, 298, 299,</u> <u>300, 302, 317, 329,</u> <u>343, 390, 401, 412,</u> <u>415, 553, 600, 614, 625,</u> 734	<u>89, 99, 162, 208,</u> <u>295, 297, 298, 328, 329,</u> <u>331, 343, 625</u>	25, <u>58, 68, 75,</u> 89, 99, 137, <u>149, 162, 176, 189, 211,</u> <u>226, 234, 239,</u> 260, 274, 295, 297, 299, 302, 317, 329, 335, <u>343, 383, 390, 401, 415,</u> 467, 496, 499, 500, 542, 550, 597, <u>625, 679</u>	<u>89, 99, 162,</u> <u>295, 297, 298,</u> <u>328, 343, 625</u>	25, 42, 48, <u>58, 68, 75,</u> 89, 99, 137, 149, 162, 175, 176, 189, 199, 200, 208, <u>211, 219, 226, 234, 239,</u> 257, 260, 274, 295, 297, 298, 299, 300, 302, 317, 328, 329, 331, 335, <u>343, 383, 390, 401, 412, 415, 456,</u> 467, 496, 499, 500, 542, 550, 553, 597, 600, 603, 614, <u>625, 679, 734</u>	<u>23/25</u>

Table 3. Cont.

Gene	M8 ^a	SLAC	FEL	REL	MEME	FUBAR	Integrated ^b	Total Common Sites ^c (X/Y)
TLR3	<u>52, 166,</u> <u>214, 237, 703</u>	<u>25, 237, 307,</u> <u>334, 703, 746</u>	<u>9, 11, 19, 25, 30, 66,</u> <u>68, 74, 214, 237, 263,</u> <u>307, 334, 346, 370, 447,</u> <u>468, 557, 698, 703,</u> <u>744, 746, 815</u>	<u>25, 30, 52, 74, 113, 137,</u> <u>158, 166, 179, 180,</u> <u>214, 237, 288, 307, 312,</u> <u>326, 334, 343, 346, 447,</u> <u>461, 463, 497, 556, 557,</u> <u>619, 703, 746, 815</u>	<u>9, 25, 30, 38, 48, 68, 74,</u> <u>93, 108, 192, 214, 234,</u> <u>237, 263, 307, 334, 346,</u> <u>349, 382, 393, 439, 447,</u> <u>451, 461, 473, 547, 557,</u> <u>577, 605, 664, 677,</u> <u>698, 703, 707, 715, 744,</u> <u>746, 815</u>	<u>25, 52, 158,</u> <u>214, 237, 334,</u> <u>346, 703, 746, 815</u>	<u>9, 11, 19, 25, 30, 38, 48, 52, 66,</u> <u>68, 74, 93, 108, 113, 137, 158, 166,</u> <u>179, 180, 192, 214, 234, 237, 263, 288,</u> <u>307, 312, 326, 334, 343, 346, 349,</u> <u>370, 382, 393, 439, 447, 451, 461,</u> <u>463, 468, 473, 497, 547, 556, 557,</u> <u>577, 605, 619, 664, 677, 698, 703,</u> <u>707, 715, 744, 746, 815</u>	<u>22/22</u>
TLR4	<u>187, 245, 270,</u> <u>271, 274, 299,</u> <u>302, 323, 352,</u> <u>370, 375, 379,</u> <u>380, 387, 398,</u> <u>403, 405, 406,</u> <u>423, 465, 522,</u> <u>595, 624,</u> <u>627, 640, 645,</u> <u>650, 655, 834</u>	<u>106, 124, 146,</u> <u>187, 245, 271,</u> <u>301, 302, 352,</u> <u>379, 380, 423,</u> <u>467, 640, 654</u>	<u>86, 95, 106, 119, 124,</u> <u>127, 141, 146, 187, 204,</u> <u>245, 270, 271, 277,</u> <u>301, 302, 329, 352, 363,</u> <u>379, 380, 403, 423, 467,</u> <u>509, 596, 603, 606, 624,</u> <u>639, 640, 654, 663, 732</u>	<u>62, 86, 106, 124, 146,</u> <u>187, 245, 270, 271, 301,</u> <u>302, 303, 323, 345, 352,</u> <u>363, 370, 379, 380, 387,</u> <u>403, 423, 430, 438, 444,</u> <u>467, 522, 596, 603, 624,</u> <u>627, 640, 654, 703</u>	<u>61, 63, 64, 85, 86, 95, 106,</u> <u>115, 119, 124, 141, 146,</u> <u>155, 180, 187, 223,</u> <u>245, 270, 271, 273, 282,</u> <u>297, 301, 302, 333, 345,</u> <u>352, 363, 370, 379, 380,</u> <u>397, 398, 403, 423, 435,</u> <u>445, 467, 469, 474, 477,</u> <u>493, 533, 548, 549, 562,</u> <u>569, 570, 596, 597, 603,</u> <u>606, 614, 624, 640, 654,</u> <u>732, 780, 784, 827</u>	<u>187, 271, 301,</u> <u>302, 352, 379,</u> <u>380, 403, 423, 467,</u> <u>522, 596, 603, 624</u>	<u>61, 62, 63, 64, 85, 86, 95,</u> <u>106, 119, 124, 127, 141, 146, 155,</u> <u>180, 187, 204, 223, 245, 270, 271,</u> <u>273, 274, 277, 282, 297, 299,</u> <u>301, 302, 303, 323, 329, 333,</u> <u>345, 352, 363, 370, 375, 379, 380,</u> <u>387, 397, 398, 403, 405, 406, 423,</u> <u>430, 435, 438, 444, 445, 465, 467,</u> <u>469, 474, 477, 493, 509, 522, 533,</u> <u>548, 549, 562, 569, 570, 595, 596,</u> <u>597, 603, 606, 614, 624, 627, 639,</u> <u>640, 645, 650, 654, 655, 663, 703,</u> <u>732, 780, 784, 827, 834</u>	<u>34/45</u>
TLR5	<u>20, 106, 130, 132,</u> <u>147, 209, 237, 281,</u> <u>468, 607, 848</u>	<u>848</u>	<u>22, 24, 35, 53, 101, 108,</u> <u>118, 130, 147, 173, 201,</u> <u>226, 231, 258, 261, 264,</u> <u>422, 466, 468, 656, 659,</u> <u>833, 848</u>	<u>20, 22, 33, 35, 106, 130,</u> <u>132, 147, 237, 258, 261,</u> <u>299, 468, 632, 646, 848</u>	<u>13, 22, 24, 87, 130, 173,</u> <u>181, 183, 196, 199, 201,</u> <u>205, 217, 226, 231, 258,</u> <u>261, 264, 265, 288, 378,</u> <u>422, 424, 466, 501, 525,</u> <u>625, 626, 632, 648, 650,</u> <u>656, 659, 679, 833, 848,</u> <u>859</u>	<u>22, 35, 130,</u> <u>261, 468, 848,</u>	<u>13, 20, 22, 24, 33, 35, 53, 87, 101,</u> <u>106, 108, 118, 130, 132, 147, 173,</u> <u>181, 183, 196, 199, 201, 205, 209, 217,</u> <u>226, 231, 237, 258, 261, 264, 265,</u> <u>281, 288, 299, 378, 422, 424, 466,</u> <u>468, 501, 525, 625, 626, 632, 646,</u> <u>648, 650, 656, 659, 679, 833, 848, 859</u>	<u>24/26</u>

Table 3. Cont.

Gene	M8 ^a	SLAC	FEL	REL	MEME	FUBAR	Integrated ^b	Total Common Sites ^c (X/Y)
TLR7	<u>65, 73, 118, 121, 122, 123, 148, 152, 156, 284, 334, 360, 395, 422, 426, 503, 524, 528, 549, 577, 677, 696, 704, 706, 722, 726, 746, 747, 758, 920</u>	<u>38, 56, 73, 121, 123, 156, 176, 395, 503, 521, 549, 577, 664, 681, 704, 722, 726, 751, 919, 920, 1049</u>	<u>56, 73, 121, 123, 156, 169, 176, 229, 253, 279, 310, 395, 503, 521, 524, 528, 549, 577, 642, 664, 681, 701, 704, 706, 722, 726, 737, 751, 851, 857, 860, 919, 920, 951, 1049</u>	<u>38, 73, 89, 95, 97, 121, 122, 123, 148, 152, 156, 176, 205, 279, 284, 334, 360, 366, 395, 398, 402, 422, 426, 494, 503, 524, 528, 549, 550, 573, 577, 678, 701, 704, 706, 712, 722, 726, 746, 920</u>	<u>38, 56, 73, 86, 95, 111, 121, 123, 156, 167, 169, 174, 199, 229, 246, 279, 310, 313, 321, 361, 377, 395, 463, 465, 495, 503, 512, 521, 524, 528, 549, 577, 622, 624, 664, 681, 686, 698, 701, 704, 706, 708, 712, 722, 726, 737, 747, 751, 768, 792, 857, 895, 919, 920, 951, 1038, 1049</u>	<u>73, 121, 123, 156, 205, 395, 503, 524, 528, 549, 577, 704, 706, 722, 726, 919, 920</u>	<u>38, 56, 65, 73, 86, 89, 95, 97, 111, 118, 121, 122, 123, 148, 152, 156, 167, 169, 174, 176, 199, 205, 229, 246, 253, 279, 284, 310, 313, 321, 334, 360, 361, 366, 377, 395, 398, 402, 422, 426, 463, 465, 494, 495, 503, 512, 521, 524, 528, 549, 550, 573, 577, 622, 624, 642, 664, 677, 678, 681, 686, 696, 698, 701, 704, 706, 708, 712, 722, 726, 737, 746, 747, 751, 758, 768, 792, 851, 857, 860, 895, 919, 920, 951, 1038, 1049</u>	<u>44/49</u>
TLR15	<u>79, 119, 185, 253, 262, 326, 333, 353, 360</u>	<u>19, 26, 33, 65, 79, 89, 114, 136, 151, 169, 191, 203, 205, 268, 289, 292, 293, 339, 343, 359, 366, 413, 436, 458, 621, 623, 627</u>	<u>26, 33, 38, 65, 89, 102, 114, 136, 145, 151, 169, 191, 203, 259, 268, 289, 292, 293, 339, 343, 359, 366, 413, 436, 458, 621, 623, 627, 649, 656, 661, 725</u>	<u>11, 13, 16, 17, 19, 26, 31, 33, 38, 48, 65, 79, 89, 102, 114, 120, 127, 128, 132, 136, 145, 149, 151, 158, 160, 164, 169, 170, 181, 191, 193, 194, 196, 200, 203, 205, 226, 249, 259, 267, 268, 289, 292, 293, 296, 315, 339, 343, 359, 363, 366, 382, 392, 413, 416, 436, 439, 450, 458, 463, 494, 495, 521, 523, 528, 550, 621, 623, 627, 649, 656, 661, 671, 676, 705, 712, 725, 813</u>	<u>12, 26, 36, 54, 65, 89, 102, 105, 107, 114, 136, 143, 151, 159, 162, 169, 175, 186, 187, 191, 203, 225, 226, 229, 230, 232, 235, 259, 263, 268, 282, 285, 289, 292, 293, 329, 335, 339, 343, 359, 363, 366, 367, 383, 400, 413, 436, 458, 485, 530, 543, 550, 621, 623, 627, 634, 651, 661, 679, 815, 839, 862, 865, 872, 873</u>	<u>26, 89, 114, 136, 169, 191, 203, 259, 268, 289, 292, 293, 339, 343, 359, 366, 413, 436, 458, 621, 623, 661</u>	<u>11, 12, 13, 16, 17, 19, 26, 31, 33, 36, 38, 48, 54, 65, 79, 89, 102, 105, 107, 114, 120, 127, 128, 132, 136, 143, 145, 149, 151, 158, 159, 160, 162, 164, 169, 170, 175, 181, 185, 186, 187, 191, 193, 194, 196, 200, 203, 205, 225, 226, 229, 230, 232, 235, 249, 259, 263, 267, 268, 282, 285, 289, 292, 293, 296, 315, 326, 329, 335, 339, 343, 359, 360, 363, 366, 367, 382, 383, 392, 400, 413, 416, 436, 439, 450, 458, 463, 485, 494, 495, 521, 523, 528, 530, 543, 550, 621, 623, 627, 634, 649, 651, 656, 661, 671, 676, 679, 705, 712, 725, 813, 815, 839, 862, 865, 872, 873</u>	<u>38/41</u>

The positively selected sites under SLAC, FEL, REL, MEME, and FUBAR were identified with default cutoff in Datamonkey online server. The sites detected as positively selected by more than one method are underlined. (a) All sites having PP > 0.90 under M8 model are shown and sites with PP > 0.99 are in bold. (b) Integrative approach includes all the sites identified by M8 (bold), SLAC, FEL, REL, MEME, and FUBAR. (c) Common sites X: Total number of sites detected by two or more methods (underlined) Y: Total number of sites detected by two or more methods plus sites with PP > 99 in M8 site model (bold).

Table 4. TreeSAAP results for positive selected sites detected with PP > 0.99 in M8 model which also had PP > 0.90 in M2a. Type I radical changes detected by TreeSAAP are underlined.

Gene	Codon	Amino Acid	M2a	M8	Total		Chemical	Structural		Other
TLR1A	408	R	2.467 +- 0.293 *	2.200 +- 0.468 **	1	1	R _F	0		0
	148	V	2.495 +- 0.123 **	1.691 +- 0.398 **	1	0		1	K ⁰	0
TLR1B	175	S	2.484 +- 0.176 *	1.688 +- 0.401 **	4	3	R _F , R _a , H _p	1	B _I	0
	298	A	2.476 +- 0.207 *	1.688 +- 0.400 **						
TLR2A	308	N	2.491 +- 0.141 **	1.691 +- 0.398 **	2	1	H	1	K ⁰	0
	16	Q	3.393 +- 0.312 **	2.819 +- 0.468 **	2	2	pH _i , H	0		0
	59	P	3.374 +- 0.374 **	2.808 +- 0.488 **	2	2	R _a , H _p	0		0
	67	<u>V</u>	<u>3.393 +- 0.313 **</u>	<u>2.819 +- 0.468 **</u>	<u>8</u>	<u>3</u>	pH _i , μ , R _a	<u>4</u>	M _V , V ⁰ , B _I , H _C	<u>1</u> <u>M_W</u>
	108	<u>G</u>	<u>3.393 +- 0.311 **</u>	<u>2.819 +- 0.467 **</u>	<u>7</u>	<u>5</u>	H, R _F , H _{nc} , H _t , R _a	<u>2</u>	<u>B_I</u>	<u>0</u>
	171	Q	3.394 +- 0.309 **	2.820 +- 0.466 **	2	2	H _t , pH _i	0		0
	206	<u>S</u>	<u>3.388 +- 0.330 **</u>	<u>2.815 +- 0.475 **</u>	<u>9</u>	<u>4</u>	R _F , μ , R _a , H _t	<u>4</u>	M _V , V ⁰ , B _I , H _C	<u>1</u> <u>M_W</u>
	304	<u>A</u>	<u>3.394 +- 0.309 **</u>	<u>2.820 +- 0.466 **</u>	<u>9</u>	<u>7</u>	R _F , H, R _a , H _p , pH _i , Pr, p	<u>2</u>	<u>K⁰, B_I</u>	<u>0</u>
	308	T	3.391 +- 0.317 **	2.818 +- 0.469 **	5	5	R _F , H, R _a , H _p , pH _i , Pr	0		0
	311	A	3.394 +- 0.309 **	2.820 +- 0.466 **	1	1	Pr	0		0
TLR2B	312	<u>R</u>	<u>3.391 +- 0.317 **</u>	<u>2.818 +- 0.469 **</u>	<u>6</u>	<u>4</u>	pH _i , H, H _{nc} , R _a	<u>2</u>	<u>V⁰, H_C</u>	<u>0</u>
	338	E	3.371 +- 0.385 **	2.807 +- 0.487 **	2	2	pH _i , R _F	0		0
	99	<u>W</u>	<u>2.510 +- 0.292 *</u>	<u>2.407 +- 0.298 **</u>	<u>6</u>	<u>5</u>	H, H _{nc} , R _a , R _F , H _t	<u>1</u>	<u>B_I</u>	<u>0</u>
	162	Q	2.524 +- 0.254 *	2.411 +- 0.289 **	5	5	pH _i , p, R _F , H _{nc} , R _a	0		0
	175	<u>E</u>	<u>2.503 +- 0.309 *</u>	<u>2.405 +- 0.303 **</u>	<u>6</u>	<u>5</u>	R _F , H, R _a , H _p , P _r	<u>1</u>	<u>K⁰</u>	<u>0</u>
	295	Q	2.506 +- 0.302 *	2.406 +- 0.301 **	2	2	pH _i , H	0		0
	456	Q	2.514 +- 0.281 *	2.409 +- 0.295 **	5	3	R _a , H _p , pH _i	2	K ⁰ , H _C	0
	214	T	2.503 +- 0.087 **	1.741 +- 0.430 **	5	5	R _a , H _p , P _r , p, pH _i	0		0
	237	R	2.500 +- 0.109 **	1.736 +- 0.435 **	3	3	R _F , H, H _{nc}	0		0
	187	S	3.497 +- 0.062 **	2.500 +- 0.019 **	3	2	pH _i , p	1	K ⁰	0
TLR3	245	N	3.497 +- 0.059 **	2.500 +- 0.017 **	1	1	P _r	0		0
	270	<u>I</u>	<u>3.498 +- 0.044 **</u>	<u>2.500 +- 0.001 **</u>	<u>7</u>	<u>4</u>	R _F , R _a , H _p , pH _i	<u>3</u>	<u>V⁰, B_I, H_C</u>	<u>0</u>
	274	<u>T</u>	<u>3.495 +- 0.102 **</u>	<u>2.499 +- 0.041 **</u>	<u>10</u>	<u>5</u>	pH _i , μ , R _a , P _r , p	<u>4</u>	<u>M_V, V⁰, B_I, H_C</u>	<u>1</u> <u>M_W</u>
	299	E	3.498 +- 0.044 **	2.500 +- 0.001 **	3	2	pH _i , P _r	1	K ⁰	0
	302	N	3.429 +- 0.411 *	2.492 +- 0.111 **	1	1	pH _i	0		0
	323	<u>N</u>	<u>3.498 +- 0.044 **</u>	<u>2.500 +- 0.004 **</u>	<u>9</u>	<u>5</u>	pH _i , μ , R _F , H, H _{nc}	<u>3</u>	<u>M_V, V⁰, H_C</u>	<u>1</u> <u>M_W</u>
	352	E	3.486 +- 0.181 **	2.498 +- 0.053 **	2	2	P _r , p	0		0
	370	D	3.496 +- 0.087 **	2.500 +- 0.028 **	3	1	pH _i	2	V ⁰ , H _C	0
	375	E	3.498 +- 0.056 **	2.500 +- 0.021 **	4	4	R _F , H _{nc} , R _a , H	0		0
	379	<u>G</u>	<u>3.498 +- 0.046 **</u>	<u>2.500 +- 0.007 **</u>	<u>6</u>	<u>5</u>	R _F , H, R _a , H _p , pH _i	<u>1</u>	<u>K⁰</u>	<u>0</u>
	380	S	3.464 +- 0.291 *	2.496 +- 0.080 **	4	2	pH _i , P _r	2	V ⁰ , H _C	0
TLR4	398	<u>I</u>	<u>3.498 +- 0.044 **</u>	<u>2.500 +- 0.001 **</u>	<u>8</u>	<u>3</u>	pH _i , μ , R _F	<u>4</u>	<u>M_V, B_I, V⁰, H_C</u>	<u>1</u> <u>M_W</u>

The ω values and Bayesian (BEB) analysis posterior probabilities are shown for sites with (**) PP > 0.99 in M8 that also have a (*) PP > 0.90 in M2a. TreeSAAP analysis results present the total number of radical changes in amino acid properties and their assigned categories. Type I sites are underlined. Properties symbols are as following: B_I: bulkiness; H: hydrophobicity; H_{nc}: normal consensus hydrophobicity; H_p: surrounding hydrophobicity; H_t: thermodynamic transfer hydrophobicity; K⁰: compressibility; μ : refractive index; M_V: molecular volume; M_W: molecular weight; P: turn tendencies; p: polarity; pH_i: isoelectric point; P_r: polar requirement; R_a: solvent accessible reduction ratio; R_F: chromatographic index; V⁰: partial specific volume; H_C: helical contact area.

4. Discussion

The vertebrate toll-like receptors (TLRs) supergene family constitutes the first line of immune defense against diverse pathogens and provides a fascinating example of the host–pathogen evolutionary arms race. Here, employing state-of-the-art DNA and protein level analyses, we provided a comprehensive comparative evolutionary genomics characterization of the vertebrate TLR supergene family using whole genome sequencing data from 79 species (mammals, birds, reptiles, amphibians, and fishes) together with strong positive selection of avian TLRs (viral and non-viral TLRs).

Our results elucidated the dynamic evolution of TLR super gene family with different level of gene gain and loss leading to variable species and lineages specific TLR repertoire across vertebrates and positive selection analysis of avian TLRs showed extensive adaptive evolution shaping host pathogen arms race in the avian TLRs.

Our results show that the TLR supergene family is broadly divided into six major families, including TLR1, TLR3, TLR4, TLR5, TLR7, and TLR11, with each having different numbers of subfamilies—e.g., 10 subfamilies in TLR1, 10 subfamilies in TLR11, 3 subfamilies in TLR7, and a single subfamily in families TLR3, TLR4, and TLR5. We conclude that most of the TLRs originated early during vertebrate evolution and that genome duplication together with differential rate of gene gain and loss shaped the TLR gene family evolution in vertebrates, leading to species and lineage-specific TLR variations. We found that gene loss was common in the avian lineage, resulting in only eight extant subfamilies, the least observed among vertebrates. In contrast, 20 subfamilies were detected in fishes, reflecting the strong impact of gene duplication. The genomic scan of diverse reptilian genomes confirmed the presence of TLR13, TLR15, TLR18, TLR21, and TLR22 in reptiles (Table 1), supporting other evidence that these TLRs are widely distributed and not limited to a particular vertebrate species or group [23,45]. For example, TLR15 was not bird-specific and is also found in most reptilian genomes. Similarly, TLR21 was also found in reptiles, as well as fishes, frogs, and birds. The subfamilies TLR11–TLR12 are mammal-specific and subfamilies TLR19–20 and TLR23–27 are specific to fishes. More duplication events occurred in fishes (TLR2, TLR4, TLR5, TLR7, TLR8, TLR9, TLR14, TLR24, and TLR19–TLR23) compared with tetrapods (TLR1, TLR2, TLR5, TLR8, TLR13, and TLR14), clearly showing a reduction of the TLR supergene family repertoire of fishes relative to other vertebrates (Table 1), e.g., maximum of 41 TLR copies in cod genome among fishes compared to maximum 14 copies in mammals suggest the greater role of TLRs in fishes. The importance of TLRs in fishes is also supported by the loss of the MHCII, CD4, and invariant chain (Ii) in cod [24] and IgM in coelacanth [28], which suggests that TLRs have an important alternative and compensatory role in the fish lineage.

The synteny of TLR supports fish- and tetrapod-specific genomic arrangements, with the exception of subfamilies TLR3, TLR5M, TLR7, and TLR8 having conserved syntenic organization across all vertebrates. The coelacanth was an exception to this trend as its TLRs had features shared with both tetrapods (TLR1, TLR2, and TLR13) and fishes (TLR21), supporting its evolutionary proximity with both groups. This pattern also probably reflects its unique immune system, which lacks IgM as part of its adaptive immune system [27,28].

The important immunological function of TLRs—i.e., protection of host from pathogens—requires them to evolve rapidly in response to selective pressure exerted by rapidly and constantly evolving pathogens. Birds have been important vectors of zoonotic pathogens and reservoirs for viruses, thus the availability of avian genomes allowed us to test the adaptive evolution of avian TLRs responsible for both viral and non-viral pathogens.

Positive selection is one of the hallmarks of immune-defense-related genes [97,98] and especially those encoding recognition proteins, evolve under positive selection [99]. There is growing evidence of positive selected sites in TLR loci [46,50,65,76,86]. These positive-selected sites can provide increased number of advantageous variations, which is important for pathogen recognition and the host–pathogen arms-race that facilitate successful adaptation against changing environments and pathogens [99]. The use of different codon and protein-level approaches revealed extensive positive selected sites in birds TLRs and the homology modeling of these important changes in TLR proteins (Tables 2–4, Figure 4 and Supplementary File 4) confirms their possible effect on structural and functional diversification of the respective TLRs, and is discussed below.

Our results were further supported by protein level approaches. Most PS sites found under high PP > 0.99 (Table 4 and Figure 4), had the highest number of positive radical physiochemical changes. Of the 71 sites with high PP > 0.99 (Table 4), 59 PS sites showed positive radical changes in physiochemical properties (TreeSAAP categories 6, 7, and 8). We also found 15 PS sites with type-I

radical changes (sites with six or more positive radical changes). TLR4 (non-viral) and TLR7 (viral) had the most sites with PP > 0.99 (Table 4). In TLR4 we found 20 such sites (Table 4 and Figure 4), among which 17 showed type II changes (sites with less than six positive radical changes) and 6 PS sites showed type I changes. Most of the type I changes were restricted to TLR2A (5 sites) and TLR4 (6 sites) (Table 4 and Figure 4). In TLR7 we found 23 sites with PP > 0.99, among which 19 showed type II changes and two PS sites showed type I changes (Table 4). The homology modeling showed that the majority of sites having PP > 0.99 were found in the ECD with only few (4 sites) found outside the ECD (Figure 4)—e.g., site 16 from TLR2A, sites 640 and 655 from TLR4, and site 920 from TLR7—were found in the signal peptide, TM domain, and TIR domain of the respective gene.

Overall positive selection in the avian TLRs suggests that the host pathogen arms race has played an important role in the rapid evolution of the avian TLRs immune genes to adapt against pathogens. Our finding suggests rapid positive selection in both viral and non-viral TLRs (Table 3) with non-viral TLR4 having the highest PS sites (20 sites with highest PP > 0.99) and among viral TLRs, TLR7 had the highest number of PS sites (23 sites with PP > 0.99). The majority of PS sites with type I and II positive radical changes were located in the LRRs of ECD, which is mainly involved in PAMPs recognition. The finding of non-viral TLR4 and viral TLR7 genes with highest number of PS sites coincides with similar finding for TLR7 [87] and non-viral TLR4 [92] in mammals. The role of non-viral TLR4 in recognizing diverse ligands (e.g., LPS-lipopolysaccharide and LTA-lipoteichoic) and TLR7 in recognizing ssRNA and possibly influenza [94] may be linked with the mechanism of how birds maintain and disseminate numerous deadly viruses [95].

These findings support the premise that the host–pathogen arms race led to co-evolution and possibly explains the strong selective pressure imposed by the long-term coexistence of viruses and birds and is consistent with the proposed role of birds as natural reservoirs and vectors of zoonotic pathogens [100]. The rapid evolution of the TLR supergene family that we have documented using newly-available bird genomes fits well with other evidence that these immune genes experienced strong adaptive selection due to the rapid evolution of pathogens [101] and that the host–pathogen arms race increases the chance of developing novel protection against diverse species of pathogens.

5. Conclusions

This study revealed lineage and species-specific differences in the distribution of the TLR supergene family among vertebrates (n = 79 species), with varied rates of gene gain/loss resulting in different repertoires of TLRs that would have facilitated recognition and protection from diverse pathogens. Our results provide strong support for the rapid evolution of both viral and non-viral avian TLRs and strengthens the hypothesis that the long-term coexistence of birds and viruses contributed to the strong selective pressure observed in the viral TLR immune genes. The patterns of gene gain, gene loss, and positive selection in the TLR gene superfamily provides compelling support of the co-evolutionary host–pathogen arms race.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1424-2818/11/8/131/s1>, Supplementary file 1: Table S1—Information of gene location used from vertebrates genomes as in Table 1, Sequences used for phylogeny of vertebrates TLRs, Avian sequences used in this study; Supplementary file 2: Synteny information of vertebrates TLRs; Supplementary file 3: Comparative domain architecture of vertebrates TLRs; Supplementary file 4: Table S2—Domain characterization of TLR1A, Table S3—Domain characterization of TLR1B, Table S4—Domain characterization of TLR2A, Table S5—Domain characterization of TLR2B, Table S6—Domain characterization of TLR3, Table S7—Domain characterization of TLR4, Table S8—Domain characterization of TLR5, Table S9—Domain characterization of TLR7, Table S10—Domain characterization of TLR15.

Author Contributions: Conceptualization, I.K. and A.A.; Methodology, I.K., E.M., and L.S.; Software, I.K. and E.M.; Validation, I.K., E.M., L.S., D.A., and A.A.; Formal analysis, I.K., E.M., L.S., and D.A.; Resources, A.A.; Data curation, I.K. and E.M.; Writing—original draft preparation, I.K.; Writing—review and editing, I.K., E.M., L.S., D.A., W.E.J., S.J.O., G.Z., E.D.J., M.T.P.G., and A.A.; Supervision and design, A.A.; Funding acquisition, A.A.

Funding: I.K., D.A. and L.S. were funded with a PhD grant from “Fundação para a Ciência e a Tecnologia” (FCT) (IK: SFRH/BD/48518/2008, DA: SFRH/BD/79766/2011, LS: SFRH/BD/134565/2017). W.E.J. prepared portions of this

manuscript while holding a National Research Council Research Associateship Award at the Walter Reed Army Institute of Research and the published material reflects the views of the authors and should not be construed to represent those of the Department of the Army or the Department of Defense. S.J.O. was supported, in part, by the Russian Science Foundation grant (project no. 17-14-01138) and by St. Petersburg State University (Genome Russia Grant no. 1.52.1647.2016). E.D.J. was supported by the Howard Hughes Medical Institute. A.A. was partially supported by the Strategic Funding UID/Multi/04423/2019 through national funds provided by FCT and the European Regional Development Fund (ERDF) in the framework of the program PT2020, by the European Structural and Investment Funds (ESIF) through the Competitiveness and Internationalization Operational Program—COMPETE 2020 and by National Funds through the FCT under the project PTDC/CTA-AMB/31774/2017 (POCI-01-0145-FEDER/031774/2017).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Akira, S.; Uematsu, S.; Takeuchi, O. Pathogen Recognition and Innate Immunity. *Cell* **2006**, *124*, 783–801. [[CrossRef](#)] [[PubMed](#)]
2. Kawai, T.; Akira, S. The role of pattern-recognition receptors in innate immunity: Update on Toll-like receptors. *Nat. Immunol.* **2010**, *11*, 373–384. [[CrossRef](#)] [[PubMed](#)]
3. Medzhitov, R. Toll-like receptors and innate immunity. *Nat. Rev. Immunol.* **2001**, *1*, 135–145. [[CrossRef](#)] [[PubMed](#)]
4. Schnare, M.; Barton, G.M.; Holt, A.C.; Takeda, K.; Akira, S.; Medzhitov, R. Toll-like receptors control activation of adaptive immune responses. *Nat. Immunol.* **2001**, *2*, 947–950. [[CrossRef](#)]
5. Iwasaki, A.; Medzhitov, R. Regulation of adaptive immunity by the innate immune system. *Science* **2010**, *327*, 291–295. [[CrossRef](#)] [[PubMed](#)]
6. Imler, J.L.; Hoffmann, J.A. Toll Receptors in Drosophila: A Family of Molecules Regulating Development and Immunity. *Curr. Top. Microbiol. Immunol.* **2002**, *270*, 63–79.
7. O'Neill, L.A.J.; Golenbock, D.; Bowie, A.G. The history of Toll-like receptors—Redefining innate immunity. *Nat. Rev. Immunol.* **2013**, *13*, 453–460. [[CrossRef](#)]
8. Janeway, C.A.J.; Medzhitov, R. Innate immune recognition. *Annu. Rev. Immunol.* **2002**, *20*, 197–216. [[CrossRef](#)]
9. Takeuchi, O.; Sato, S.; Horiuchi, T.; Hoshino, K.; Takeda, K.; Dong, Z.; Modlin, R.L.; Akira, S. Cutting edge: Role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. *J. Immunol.* **2002**, *169*, 10–14. [[CrossRef](#)]
10. Gantner, B.N.; Simmons, R.M.; Canavera, S.J.; Akira, S.; Underhill, D.M. Collaborative induction of inflammatory responses by dectin-1 and Toll-like receptor 2. *J. Exp. Med.* **2003**, *197*, 1107–1117. [[CrossRef](#)]
11. Kawai, T.; Akira, S. TLR signaling. *Cell Death Differ.* **2006**, *13*, 816–825. [[CrossRef](#)] [[PubMed](#)]
12. Brodsky, I.; Medzhitov, R. Two Modes of Ligand Recognition by TLRs. *Cell* **2007**, *130*, 979–981. [[CrossRef](#)] [[PubMed](#)]
13. Lee, J.; Jin, M.; Kim, S.; Heo, J. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* **2007**, *130*, 1071–1082.
14. Jin, M.S.; Lee, J.O. Structures of TLR-ligand complexes. *Curr. Opin. Immunol.* **2008**, *20*, 414–419. [[CrossRef](#)] [[PubMed](#)]
15. Botos, I.; Segal, D.M.; Davies, D.R. The Structural Biology of Toll-Like Receptors. *Structure* **2011**, *19*, 447–459. [[CrossRef](#)] [[PubMed](#)]
16. O'Neill, L.A.J.; Bowie, A.G. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat. Rev. Immunol.* **2007**, *7*, 353–364. [[CrossRef](#)] [[PubMed](#)]
17. Doyle, S.L.; O'Neill, L.A.J. Toll-like receptors: From the discovery of NFκB to new insights into transcriptional regulations in innate immunity. *Biochem. Pharmacol.* **2006**, *72*, 1102–1113. [[CrossRef](#)] [[PubMed](#)]
18. Takeda, K.; Akira, S. Microbial recognition by Toll-like receptors. *J. Dermatol. Sci.* **2004**, *34*, 73–82. [[CrossRef](#)] [[PubMed](#)]
19. Medzhitov, R. Recognition of microorganisms and activation of the immune response. *Nature* **2007**, *449*, 819–826. [[CrossRef](#)] [[PubMed](#)]

20. Kumar, H.; Kawai, T.; Akira, S. Toll-like receptors and innate immunity. *Biochem. Biophys. Res. Commun.* **2009**, *388*, 621–625. [[CrossRef](#)] [[PubMed](#)]
21. Medzhitov, R.; Preston-Hurlburt, P.; Janeway, C.A. A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature* **1997**, *388*, 394–397. [[CrossRef](#)] [[PubMed](#)]
22. Matsushima, N.; Tanaka, T.; Enkhbayar, P.; Mikami, T.; Taga, M.; Yamada, K.; Kuroki, Y. Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genom.* **2007**, *8*, 124. [[CrossRef](#)] [[PubMed](#)]
23. Roach, J.C.; Glusman, G.; Rowen, L.; Kaur, A.; Purcell, M.K.; Smith, K.D.; Hood, L.E.; Aderem, A. The evolution of vertebrate Toll-like receptors. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 9577–9582. [[CrossRef](#)] [[PubMed](#)]
24. Star, B.; Nederbragt, A.J.; Jentoft, S.; Grimholt, U.; Malmstrøm, M.; Gregers, T.F.; Rounge, T.B.; Paulsen, J.; Solbakken, M.H.; Sharma, A.; et al. The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **2011**, *477*, 207–210. [[CrossRef](#)] [[PubMed](#)]
25. Flajnik, M.F.; Kasahara, M. Origin and evolution of the adaptive immune system: Genetic events and selective pressures. *Nat. Rev. Genet.* **2010**, *11*, 47–59. [[CrossRef](#)] [[PubMed](#)]
26. Litman, G.W.; Rast, J.P.; Fugmann, S.D. The origins of vertebrate adaptive immunity. *Nat. Rev. Immunol.* **2010**, *10*, 543–553. [[CrossRef](#)] [[PubMed](#)]
27. Boudinot, P.; Zou, J.; Ota, T.; Buonocore, F.; Scapigliati, G.; Canapa, A.; Cannon, J.; Litman, G.; Hansen, J.D. A tetrapod-like repertoire of innate immune receptors and effectors for coelacanths. *J. Exp. Zool. Part B Mol. Dev. Evol.* **2014**, *322*, 415–437. [[CrossRef](#)] [[PubMed](#)]
28. Amemiya, C.T.; Alföldi, J.; Lee, A.P.; Fan, S.; Philippe, H.; MacCallum, I.; Braasch, I.; Manousaki, T.; Schneider, I.; Rohner, N.; et al. The African coelacanth genome provides insights into tetrapod evolution. *Nature* **2013**, *496*, 311–316. [[CrossRef](#)]
29. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
30. Jarvis, E.D.; Mirarab, S.; Aberer, A.J.; Li, B.; Houde, P.; Li, C.; Ho, S.Y.W.; Faircloth, B.C.; Nabholz, B.; Howard, J.T.; et al. Whole Genome Analyses Resolve Early Branches in the Tree of Life of Modern Birds. *Science* **2014**, *346*, 1320–1331. [[CrossRef](#)]
31. Zhang, G.; Li, C.; Li, Q.; Li, B.; Larkin, D.M.; Lee, C.; Storz, J.F.; Antunes, A.; Greenwold, M.J.; Meredith, R.W.; et al. Comparative Genomics Reveals Insights into Avian Genome Evolution and Adaptation. *Science* **2014**, *346*, 1311–1320. [[CrossRef](#)] [[PubMed](#)]
32. Muffato, M.; Louis, A.; Poisnel, C.E.; Crollius, H.R. Genomicus: A database and a browser to study gene synteny in modern and ancestral genomes. *Bioinformatics* **2010**, *26*, 1119–1121. [[CrossRef](#)]
33. Louis, A.; Muffato, M.; Roest Crollius, H. Genomicus: Five genome browsers for comparative genomics in eukaryota. *Nucleic Acids Res.* **2012**, *41*, D700–D705. [[CrossRef](#)] [[PubMed](#)]
34. Pereira, S.; Vasconcelos, V.M.; Antunes, A. The phosphoprotein phosphatase family of Ser/Thr phosphatases as principal targets of naturally occurring toxins. *Crit. Rev. Toxicol.* **2011**, *41*, 83–110. [[CrossRef](#)] [[PubMed](#)]
35. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [[CrossRef](#)] [[PubMed](#)]
36. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **2004**, *5*, 113. [[CrossRef](#)] [[PubMed](#)]
37. Gouy, M.; Guindon, S.; Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **2010**, *27*, 221–224. [[CrossRef](#)]
38. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [[CrossRef](#)]
39. Guindon, S.; Gascuel, O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.* **2003**, *52*, 696–704. [[CrossRef](#)]
40. Felsenstein, J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* **1985**, *39*, 783–791. [[CrossRef](#)]
41. Xia, X. DAMBE5: A Comprehensive Software Package for Data Analysis in Molecular Biology and Evolution. *Mol. Biol. Evol.* **2013**, *30*, 1720–1728. [[CrossRef](#)] [[PubMed](#)]

42. Huelsenbeck, J.P.; Rannala, B. Phylogenetic Methods Come of Age: Testing Hypotheses in an Evolutionary Context. *Science* **1997**, *276*, 227–232. [[CrossRef](#)] [[PubMed](#)]
43. Xia, X.; Xie, Z.; Salemi, M.; Chen, L.; Wang, Y. An index of substitution saturation and its application. *Mol. Phylogenet. Evol.* **2003**, *26*, 1–7. [[CrossRef](#)]
44. Pond, S.L.K.; Posada, D.; Gravenor, M.B.; Woelk, C.H.; Frost, S.D.W. GARD: A genetic algorithm for recombination detection. *Bioinformatics* **2006**, *22*, 3096–3098. [[CrossRef](#)] [[PubMed](#)]
45. Sawyer, S. Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **1989**, *6*, 526–538.
46. Alcaide, M.; Edwards, S.V. Molecular Evolution of the Toll-Like Receptor Multigene Family in Birds. *Mol. Biol. Evol.* **2011**, *28*, 1703–1715. [[CrossRef](#)] [[PubMed](#)]
47. Pereira, S.R.; Vasconcelos, V.; Antunes, A. Computational Study of the Covalent Bonding of Microcystins to Cysteine Residues—A Reaction Involved in the Inhibition of the PPP Family of Protein Phosphatases. *FEBS J.* **2013**, *280*, 674–680. [[CrossRef](#)] [[PubMed](#)]
48. Cormican, P.; Lloyd, A.T.; Downing, T.; Connell, S.J.; Bradley, D.; O’Farrelly, C. The avian Toll-Like receptor pathway—Subtle differences amidst general conformity. *Dev. Comp. Immunol.* **2009**, *33*, 967–973. [[CrossRef](#)]
49. Grueber, C.E.; Wallis, G.P.; King, T.M.; Jamieson, I.G. Variation at Innate Immunity Toll-Like Receptor Genes in a Bottlenecked Population of a New Zealand Robin. *PLoS ONE* **2012**, *7*, e45011. [[CrossRef](#)]
50. Velová, H.; Gutowska-Ding, M.W.; Burt, D.W.; Vinkler, M. Toll-like receptor evolution in birds: Gene duplication, pseudogenisation and diversifying selection. *Mol. Biol. Evol.* **2018**, *35*, 2170–2184. [[CrossRef](#)]
51. Yang, Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **1997**, *13*, 555–556. [[CrossRef](#)]
52. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)] [[PubMed](#)]
53. Maldonado, E.; Almeida, D.; Escalona, T.; Khan, I.; Vasconcelos, V.; Antunes, A. LMAP: Lightweight Multigene Analyses in PAML. *BMC Bioinform.* **2016**, *17*, 354. [[CrossRef](#)] [[PubMed](#)]
54. Pond, S.L.K.; Frost, S.D.W.; Pond, S.L.K. Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **2005**, *21*, 2531–2533. [[CrossRef](#)] [[PubMed](#)]
55. Woolley, S.; Johnson, J.; Smith, M.J.; Crandall, K.A.; McClellan, D.A. TreeSAAP: Selection on Amino Acid Properties using phylogenetic trees. *Bioinformatics* **2003**, *19*, 671–672. [[CrossRef](#)] [[PubMed](#)]
56. Pond, S.L.K.; Muse, S.V. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **2005**, *21*, 676–679. [[CrossRef](#)] [[PubMed](#)]
57. Delpont, W.; Poon, A.F.Y.; Frost, S.D.W.; Kosakovsky Pond, S.L. Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **2010**, *26*, 2455–2457. [[CrossRef](#)] [[PubMed](#)]
58. Pond, S.L.K.; Frost, S.D.W. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol. Biol. Evol.* **2005**, *22*, 1208–1222. [[CrossRef](#)] [[PubMed](#)]
59. Murrell, B.; Wertheim, J.O.; Moola, S.; Weighill, T.; Scheffler, K.; Pond, S.L.K. Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genet.* **2012**, *8*, e1002764. [[CrossRef](#)]
60. Murrell, B.; Moola, S.; Mabona, A.; Weighill, T.; Sheward, D.; Pond, S.L.K.; Scheffler, K. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Mol. Biol. Evol.* **2013**, *30*, 1196–1205. [[CrossRef](#)]
61. Fry, B.G.; Undheim, E.A.B.; Ali, S.A.; Debono, J.; Scheib, H.; Ruder, T.; Jackson, T.N.W.; Morgenstern, D.; Cadwallader, L.; Whitehead, D.; et al. Squeezers and leaf-cutters: Differential diversification and degeneration of the venom system in toxicoferan reptiles. *Mol. Cell. Proteom.* **2013**, *12*, 1881–1899. [[CrossRef](#)]
62. Undheim, E.A.B.; Sunagar, K.; Herzig, V.; Kely, L.; Low, D.H.W.; Jackson, T.N.W.; Jones, A.; Kurniawan, N.; King, G.F.; Ali, S.A.; et al. A Proteomics and Transcriptomics Investigation of the Venom from the Barychelid Spider *Tritrame loki* (Brush-Foot Trapdoor). *Toxins* **2013**, *5*, 2488–2503. [[CrossRef](#)] [[PubMed](#)]
63. Sunagar, K.; Fry, B.; Jackson, T.N.W.; Casewell, N.R.; Undheim, E.A.B.; Vidal, N.; Ali, S.A.; King, G.F.; Vasudevan, K.; Vasconcelos, V.; et al. Molecular evolution of vertebrate neurotrophins: Co-option of the highly conserved nerve growth factor gene into the advanced snake venom arsenal. *PLoS ONE* **2013**, *8*, e81827. [[CrossRef](#)]
64. Temperley, N.D.; Berlin, S.; Paton, I.R.; Griffin, D.K.; Burt, D.W. Evolution of the chicken Toll-like receptor gene family: A story of gene gain and gene loss. *BMC Genom.* **2008**, *9*, 62. [[CrossRef](#)] [[PubMed](#)]
65. Sundaram, A.Y.; Kiron, V.; Dopazo, J.; Fernandes, J.M. Diversification of the expanded teleost-specific toll-like receptor family in Atlantic cod, *Gadus morhua*. *BMC Evol. Biol.* **2012**, *12*, 256. [[CrossRef](#)] [[PubMed](#)]

66. Blomme, T.; Vandepoele, K.; De Bodt, S.; Simillion, C.; Maere, S.; Van De Peer, Y. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol.* **2006**, *7*, R43. [[CrossRef](#)] [[PubMed](#)]
67. DeMuth, J.P.; De Bie, T.; Stajich, J.E.; Cristianini, N.; Hahn, M.W. The Evolution of Mammalian Gene Families. *PLoS ONE* **2006**, *1*, e85. [[CrossRef](#)] [[PubMed](#)]
68. Almeida, D.; Maldonado, E.; Khan, I.; Silva, L.; Gilbert, M.T.P.; Zhang, G.; Jarvis, E.D.; O'Brien, S.J.; Johnson, W.E.; Antunes, A. Whole-Genome Identification, Phylogeny, and Evolution of the Cytochrome P450 Family 2 (CYP2) Subfamilies in Birds. *Genome Biol. Evol.* **2016**, *8*, 1115–1131. [[CrossRef](#)] [[PubMed](#)]
69. da Fonseca, R.; Antunes, A.; Melo, A.; Ramos, M.J. Structural divergence and adaptive evolution in mammalian cytochromes P450 2C. *Gene* **2007**, *387*, 58–66. [[CrossRef](#)]
70. Machado, J.P.; Johnson, E.W.; O'Brien, S.J.; Vasconcelos, V.; Antunes, A. Adaptive evolution of the matrix extracellular phosphoglycoprotein in mammals. *BMC Evol. Biol.* **2011**, *11*, 342. [[CrossRef](#)]
71. Khan, I.; Maldonado, E.; Vasconcelos, V.; O'Brien, S.J.; Johnson, W.E.; Antunes, A. Mammalian keratin associated proteins (KRTAPs) subgenomes: Disentangling hair diversity and adaptation to terrestrial and aquatic environments. *BMC Genom.* **2014**, *15*, 779. [[CrossRef](#)]
72. Khan, I.; Yang, Z.; Maldonado, E.; Li, C.; Zhang, G.; Gilbert, M.T.P.; Jarvis, E.D.; O'Brien, S.J.; Johnson, W.E.; Antunes, A. Olfactory Receptor Subgenomes Linked with Broad Ecological Adaptations in Sauropsida. *Mol. Biol. Evol.* **2015**, *32*, 2832–2843. [[CrossRef](#)]
73. Silva, L.; Antunes, A. Vomeronasal Receptors in Vertebrates and the Evolution of Pheromone Detection. *Annu. Rev. Anim. Biosci.* **2017**, *5*, 353–370. [[CrossRef](#)]
74. Luo, S.J.; Johnson, W.E.; Martenson, J.; Antunes, A.; Martelli, P.; Uphyrkina, O.; Traylor-Holzer, K.; Smith, J.L.; O'Brien, S.J. Subspecies Genetic Assignments of Worldwide Captive Tigers Increase Conservation Value of Captive Populations. *Curr. Biol.* **2008**, *18*, 592–596. [[CrossRef](#)]
75. Kasamatsu, J.; Oshiumi, H.; Matsumoto, M.; Kasahara, M.; Seya, T. Phylogenetic and expression analysis of lamprey toll-like receptors. *Dev. Comp. Immunol.* **2010**, *34*, 855–865. [[CrossRef](#)]
76. Zhang, J.; Liu, S.; Rajendran, K.; Sun, L.; Zhang, Y.; Sun, F.; Kucuktas, H.; Liu, H.; Liu, Z. Pathogen recognition receptors in channel catfish: III Phylogeny and expression analysis of Toll-like receptors. *Dev. Comp. Immunol.* **2013**, *40*, 185–194. [[CrossRef](#)]
77. Huang, Y.; Temperley, N.D.; Ren, L.; Smith, J.; Li, N.; Burt, D.W. Molecular evolution of the vertebrate TLR1 gene family—A complex history of gene duplication, gene conversion, positive selection and co-evolution. *BMC Evol. Biol.* **2011**, *11*, 149. [[CrossRef](#)]
78. Kang, J.Y.; Lee, J.O. Structural Biology of the Toll-Like Receptor Family. *Annu. Rev. Biochem.* **2011**, *80*, 917–941. [[CrossRef](#)]
79. Kang, J.Y.; Nan, X.; Jin, M.S.; Youn, S.J.; Ryu, Y.H.; Mah, S.; Han, S.H.; Lee, H.; Paik, S.G.; Lee, J.O. Recognition of Lipopeptide Patterns by Toll-like Receptor 2–Toll-like Receptor 6 Heterodimer. *Immunity* **2009**, *31*, 873–884. [[CrossRef](#)]
80. Guan, Y.; Ranoa, D.R.E.; Jiang, S.; Mutha, S.K.; Li, X.; Baudry, J.; Tapping, R.I. Human TLRs 10 and 1 share common mechanisms of innate immune sensing but not signaling. *J. Immunol.* **2010**, *184*, 5094–5103. [[CrossRef](#)]
81. Wlasiuk, G.; Nachman, M.W. Adaptation and Constraint at Toll-Like Receptors in Primates. *Mol. Biol. Evol.* **2010**, *27*, 2172–2186. [[CrossRef](#)]
82. Liu, L.; Botos, I.; Wang, Y.; Leonard, J.N.; Shiloach, J.; Segal, D.M.; Davies, D.R. Structural basis of Toll-Like Receptor 3 Signaling with double-stranded RNA. *Science* **2008**, *320*, 379–381. [[CrossRef](#)]
83. Kim, H.M.; Park, B.S.; Kim, J.I.; Kim, S.E.; Lee, J.; Oh, S.C.; Enkhbayar, P.; Matsushima, N.; Lee, H.; Yoo, O.J.; et al. Crystal Structure of the TLR4-MD-2 Complex with Bound Endotoxin Antagonist Eritoran. *Cell* **2007**, *130*, 906–917. [[CrossRef](#)]
84. Park, B.S.; Song, D.H.; Kim, H.M.; Choi, B.S.; Lee, H.; Lee, J.O. The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* **2009**, *458*, 1191–1195. [[CrossRef](#)]
85. Kumar, H.; Kawai, T.; Akira, S. Pathogen recognition in the innate immune response. *Biochem. J.* **2009**, *420*, 1–16. [[CrossRef](#)]
86. Hajjar, A.M.; Ernst, R.K.; Tsai, J.H.; Wilson, C.B.; Miller, S.I. Human Toll-like receptor 4 recognizes host-specific LPS modifications. *Nat. Immunol.* **2002**, *3*, 354–359. [[CrossRef](#)]

87. Hayashi, F.; Smith, K.D.; Ozinsky, A.; Hawn, T.R.; Yi, E.C.; Goodlett, D.R.; Eng, J.K.; Akira, S.; Underhill, D.M.; Aderem, A. The innate immune response to bacterial flagellin is mediated by Toll-like receptor. *Nature* **2001**, *410*, 1099–1103. [[CrossRef](#)]
88. Andersen-Nissen, E.; Smith, K.D.; Bonneau, R.; Strong, R.K.; Aderem, A. A conserved surface on Toll-like receptor 5 recognizes bacterial flagellin. *J. Exp. Med.* **2007**, *204*, 393–403. [[CrossRef](#)]
89. Andersen-Nissen, E.; Smith, K.D.; Strobe, K.L.; Barrett, S.L.R.; Cookson, B.T.; Logan, S.M.; Aderem, A. Evasion of Toll-like receptor 5 by flagellated bacteria. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 9247–9252. [[CrossRef](#)]
90. Gantier, M.P.; Tong, S.; Behlke, M.A.; Xu, D.; Phipps, S.; Foster, P.S.; Williams, B.R.G. TLR7 is involved in sequence-specific sensing of single-stranded RNAs in human macrophages. *J. Immunol.* **2008**, *180*, 2117–2124. [[CrossRef](#)]
91. Wang, L.F.; Walker, P.J.; Poon, L.L. Mass extinctions, biodiversity and mitochondrial function: Are bats 'special' as reservoirs for emerging viruses? *Curr. Opin. Virol.* **2011**, *1*, 649–657. [[CrossRef](#)]
92. Zhang, G.; Cowled, C.; Shi, Z.; Huang, Z.; Bishop-Lilly, K.A.; Fang, X.; Wynne, J.W.; Xiong, Z.; Baker, M.L.; Zhao, W.; et al. Comparative Analysis of Bat Genomes Provides Insight into the Evolution of Flight and Immunity. *Science* **2013**, *339*, 456–460. [[CrossRef](#)]
93. Lund, J.M.; Alexopoulou, L.; Sato, A.; Karow, M.; Adams, N.C.; Gale, N.W.; Iwasaki, A.; Flavell, R.A. Recognition of single-stranded RNA viruses by Toll-like receptor. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5598–5603. [[CrossRef](#)]
94. Boyd, A.C.; Peroval, M.Y.; Hammond, J.A.; Prickett, M.D.; Young, J.R.; Smith, A.L. TLR15 Is Unique to Avian and Reptilian Lineages and Recognizes a Yeast-Derived Agonist. *J. Immunol.* **2012**, *189*, 4930–4938. [[CrossRef](#)]
95. Areal, H.; Abrantes, J.; Esteves, P.J. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol. Biol.* **2011**, *11*, 368. [[CrossRef](#)]
96. Slack, J.L. Identification of Two Major Sites in the Type I Interleukin-1 Receptor Cytoplasmic Region Responsible for Coupling to Pro-inflammatory Signaling Pathways. *J. Biol. Chem.* **2000**, *275*, 4670–4678. [[CrossRef](#)]
97. Lahn, B.T.; Vallender, E.J. Positive selection on the human genome. *Hum. Mol. Genet.* **2004**, *13* (Suppl. S2), R245–R254.
98. Sackton, T.B.; Lazzaro, B.P.; A Schlenke, T.; Evans, J.D.; Hultmark, D.; Clark, A.G. Dynamic evolution of the innate immune system in *Drosophila*. *Nat. Genet.* **2007**, *39*, 1461–1468. [[CrossRef](#)]
99. Kosiol, C.; Vinar, T.; Da Fonseca, R.R.; Hubisz, M.J.; Bustamante, C.D.; Nielsen, R.; Siepel, A. Patterns of Positive Selection in Six Mammalian Genomes. *PLoS Genet.* **2008**, *4*, e1000144. [[CrossRef](#)]
100. Reed, K.D.; Meece, J.K.; Henkel, J.S.; Shukla, S.K. Birds, Migration and Emerging Zoonoses: West Nile Virus, Lyme Disease, Influenza A and Enteropathogens. *Clin. Med. Res.* **2003**, *1*, 5–12. [[CrossRef](#)]
101. Nielsen, R.; Bustamante, C.; Clark, A.G.; Glanowski, S.; Sackton, T.B.; Hubisz, M.J.; Fledel-Alon, A.; Tanenbaum, D.M.; Civello, D.; White, T.J.; et al. A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees. *PLoS Biol.* **2005**, *3*, e170. [[CrossRef](#)]

